



## **BTEP Presentation: Variant Analysis**

**Keyur Talsania**

Advanced Biomedical and Computational Science (ABCS)

Bioinformatics and Computational Science (BACS) Directorate

Frederick National Laboratory for Cancer Research

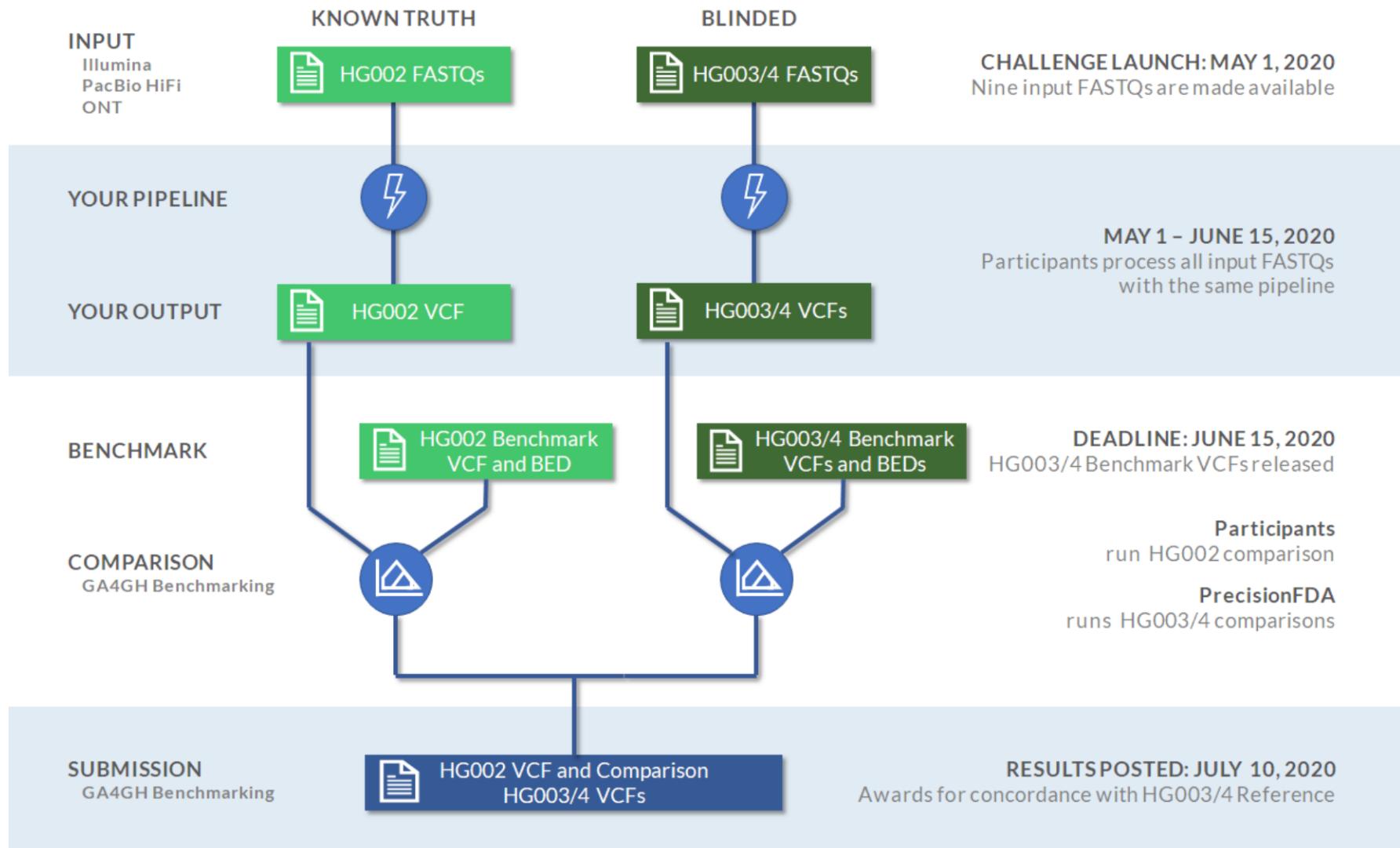
# *Precision FDA Challenge V2*

*Calling Variants from Short and Long Reads in Difficult-to-Map Regions*

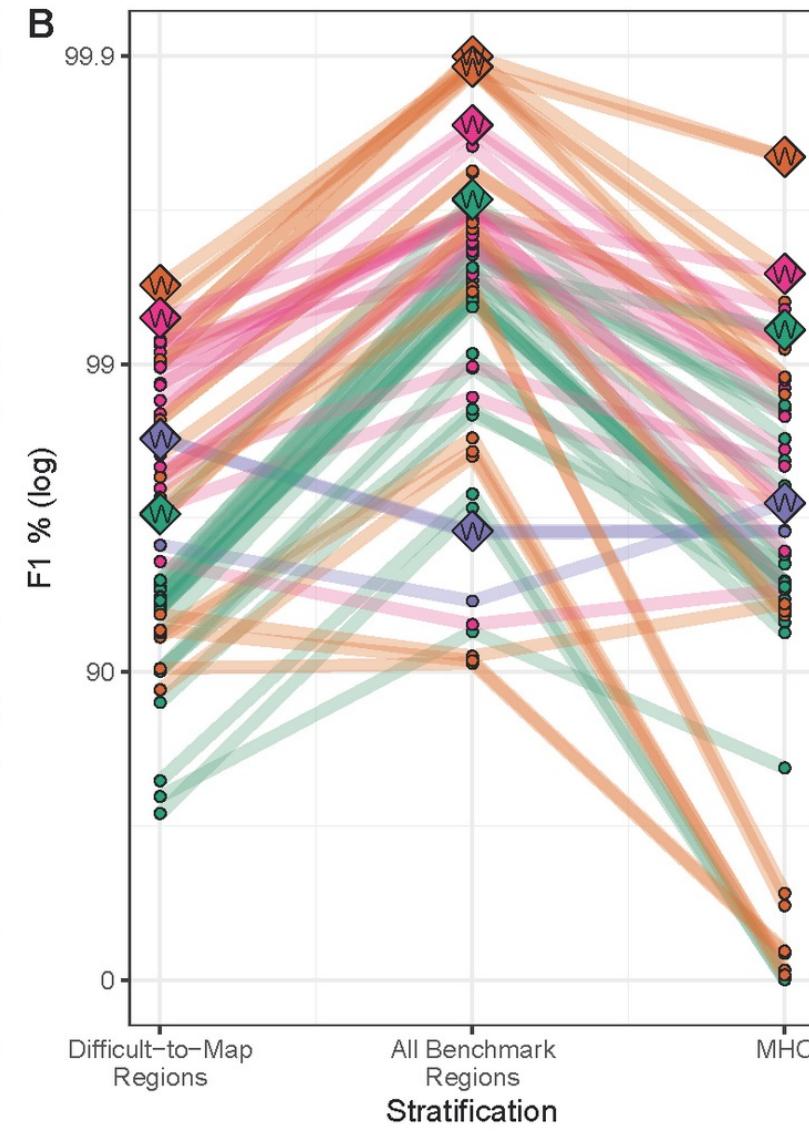
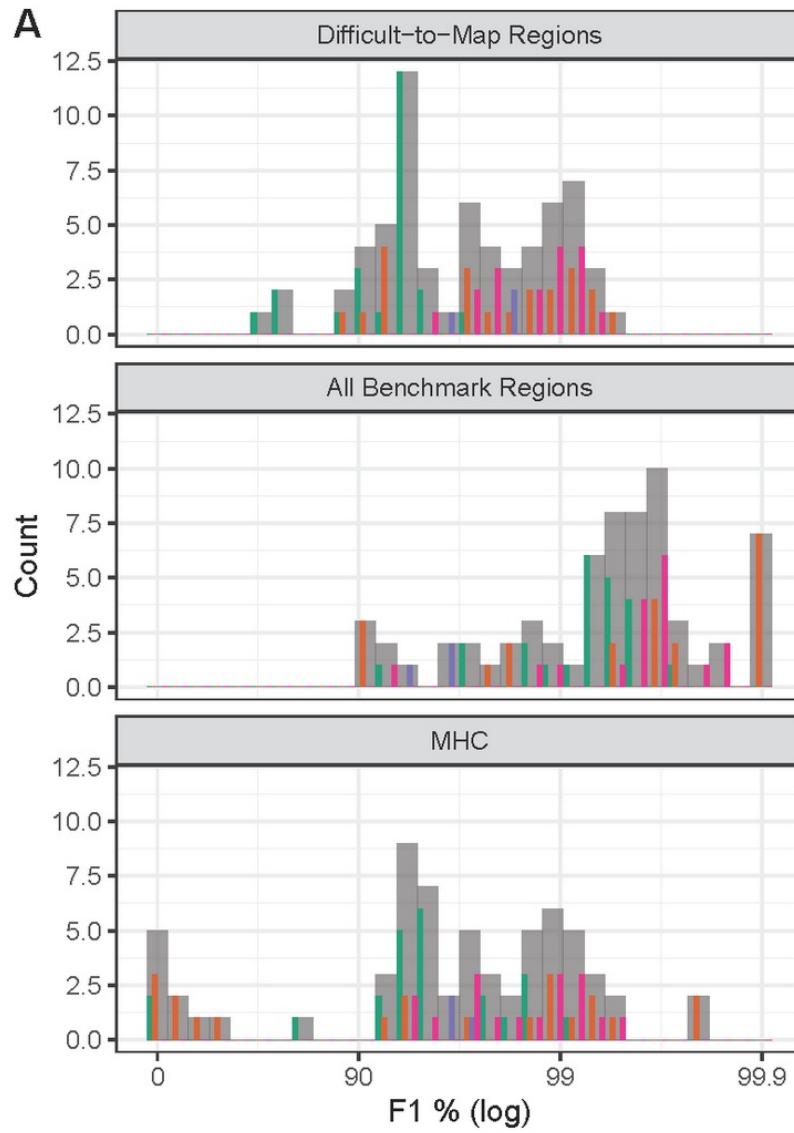
## Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions

- 9 Precision FDA-provided input datasets, specifically WGS from **Illumina (~35x), PacBio HiFi (~35x), and Oxford Nanopore (~50-80x)** of the HG002 (NA24385), HG003 (NA24149), and HG004 (NA24143) human samples.
- Variant calling in challenging genomic regions in GRCh38
  - Segment Dups, Major Histocompatibility Complex (MHC)
- HG002 truth set was provided.

# Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions



# Truth Challenge V2: Results



Technology    ILLUMINA    MULTI    ONT    PACBIO

# Truth Challenge V2: Results - Illumina

BEST PERFORMANCE

**Difficult-to-Map Regions – ILLUMINA**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**DRAGEN team at Illumina**

DRAGEN Experimental Extension into Difficult  
Regions - Illumina reads

BEST PERFORMANCE

**All Benchmark Regions – ILLUMINA**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**DRAGEN team at Illumina**

DRAGEN Experimental Extension into  
Difficult Regions - Illumina reads

BEST PERFORMANCE

**MHC – ILLUMINA**

in the precisionFDA Truth Challenge V2



AWARDED TO

**Seven Bridges Genomics**

Seven Bridges GRAF - Illumina

# Truth Challenge V2: Results - Pacbio

BEST PERFORMANCE  
**Difficult-to-Map Regions – PACBIO**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**Sentieon**

PacBio HIFI only submission

BEST PERFORMANCE  
**All Benchmark Regions – PACBIO**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**The Genomics Team in Google Health**

DeepVariant PacBio

BEST PERFORMANCE  
**MHC – PACBIO**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**Sentieon**

PacBio HIFI only submission

# Truth Challenge V2: Results - ONT

BEST PERFORMANCE

**Difficult-to-Map Regions – ONT**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**UCSC CGL and Google Health**

PEPPER-DeepVariant (ONT\_9to6)

BEST PERFORMANCE

**All Benchmark Regions – ONT**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**UCSC CGL and Google Health**

PEPPER-DeepVariant (ONT\_9to6)

BEST PERFORMANCE

**MHC – ONT**  
in the precisionFDA Truth Challenge V2



AWARDED TO

**Wang Genomics Lab**

ONT NanoCaller Clair Medaka Ensemble

# Truth Challenge V2: Results – Multi

BEST PERFORMANCE

Difficult-to-Map Regions – MULTI  
in the precisionFDA Truth ChallengeV2



AWARDED TO

**Roche Sequencing Solutions**

RN-Illumina-PacBio-ONT

BEST PERFORMANCE

All Benchmark Regions – MULTI  
in the precisionFDA Truth ChallengeV2



AWARDED TO

**Sentieon**

Combination of Illumina, PacBio HIFI, and Oxford  
Nanopore submission



BEST PERFORMANCE

All Benchmark Regions – MULTI  
in the precisionFDA Truth ChallengeV2



AWARDED TO

**Roche Sequencing Solutions**

RN-Illumina-PacBio-ONT



BEST PERFORMANCE

All Benchmark Regions – MULTI  
in the precisionFDA Truth ChallengeV2



AWARDED TO

**The Genomics Team in Google Health**

DeepVariant Hybrid PacBio-Illumina

BEST PERFORMANCE

MHC – MULTI

in the precisionFDA Truth ChallengeV2



AWARDED TO

**Sentieon**

Combination of Illumina, PacBio HIFI, and Oxford  
Nanopore submission

# Fastest Sequencing

[RECORDS](#)[SPOTLIGHT](#)[BOOKS](#)[BUSINESS SOLUTIONS](#)[NEWS](#)[ABOUT US](#)

EN



## "A new frontier": Fastest ever DNA sequencing technique achieved

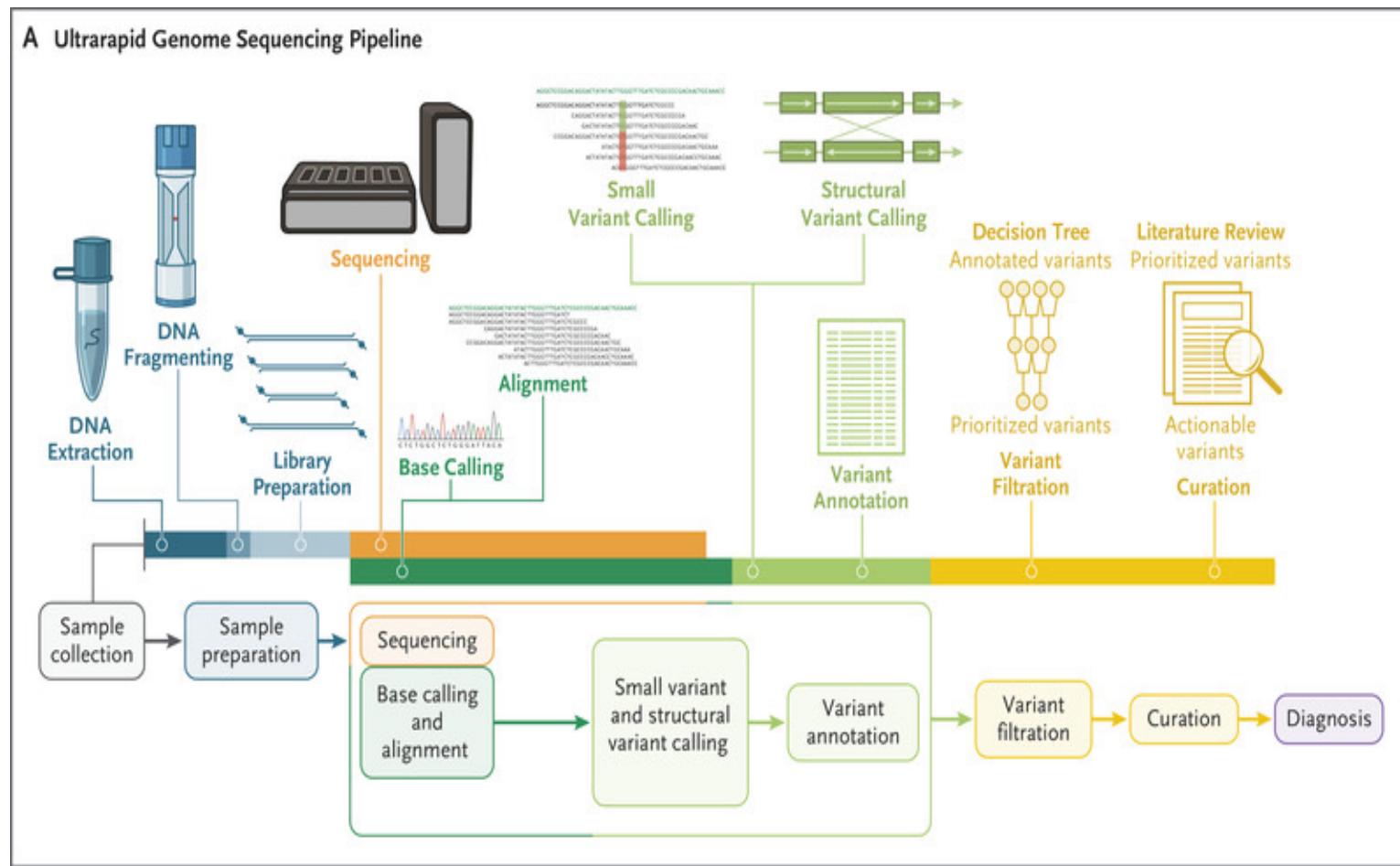
By Ana Rahlves | Published 05 May 2022

### Fastest DNA sequencing technique helps undiagnosed patients find answers in mere hours

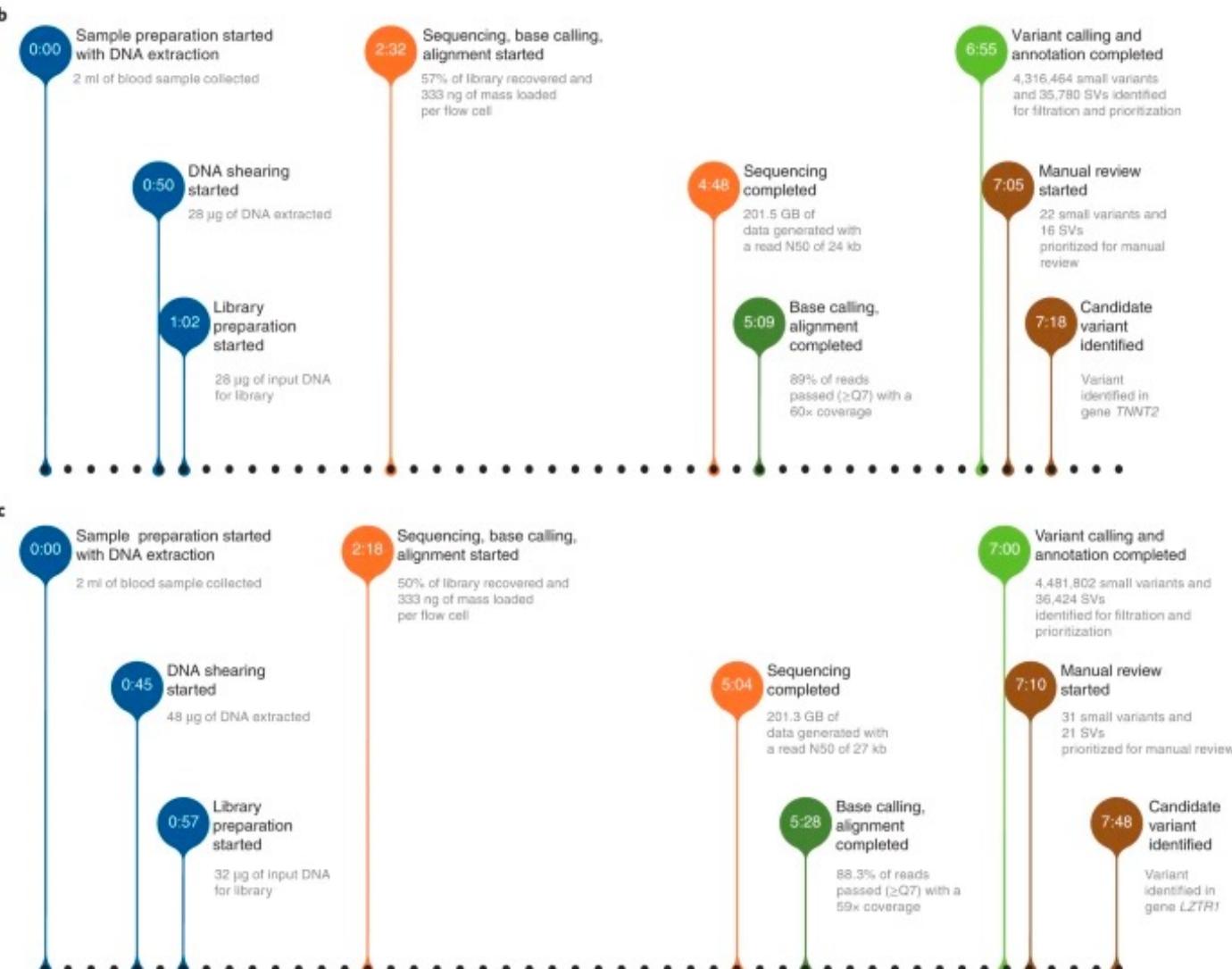
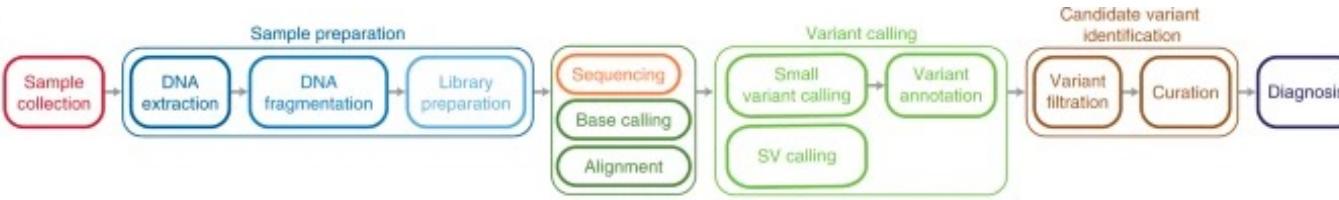
A research effort led by Stanford scientists set the first Guinness World Record for the fastest DNA sequencing technique, which was used to sequence a human genome in just 5 hours and 2 minutes.

The fastest run time was 7 hours 18 minutes!

# Fastest Sequencing



# Fastest Sequencing



*In house Precision FDA analysis*

# Tools used for comparison

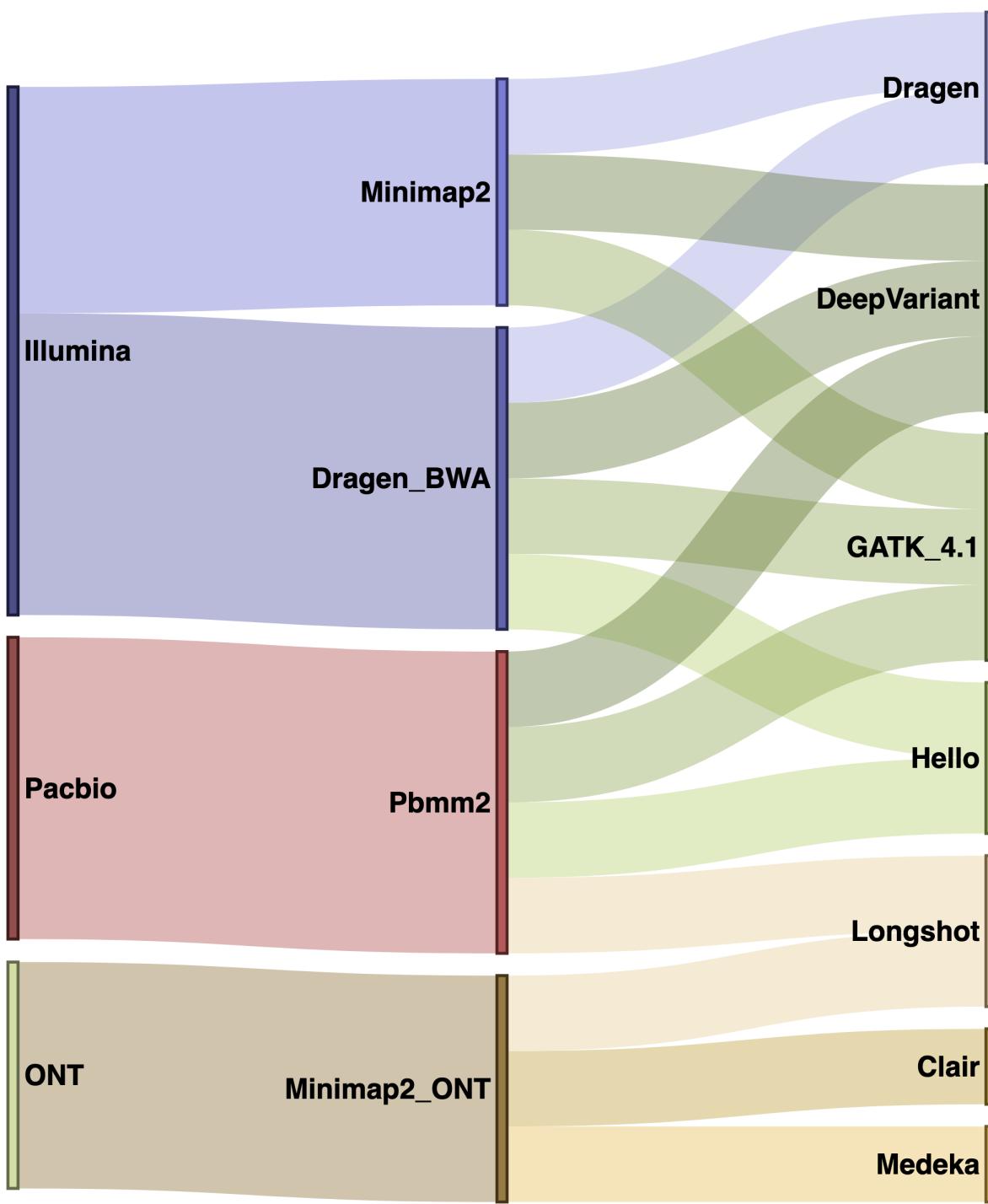
| Tools                     | Version tested |
|---------------------------|----------------|
| Deepvariant               | 3              |
| Clair / Clair3            | 2              |
| Longshot                  | 1              |
| PEPPER-Margin-DeepVariant | 1              |
| Dragen*                   | 4              |
| Octopus                   | 2              |
| GATK                      | 2              |

# Tools used for comparison

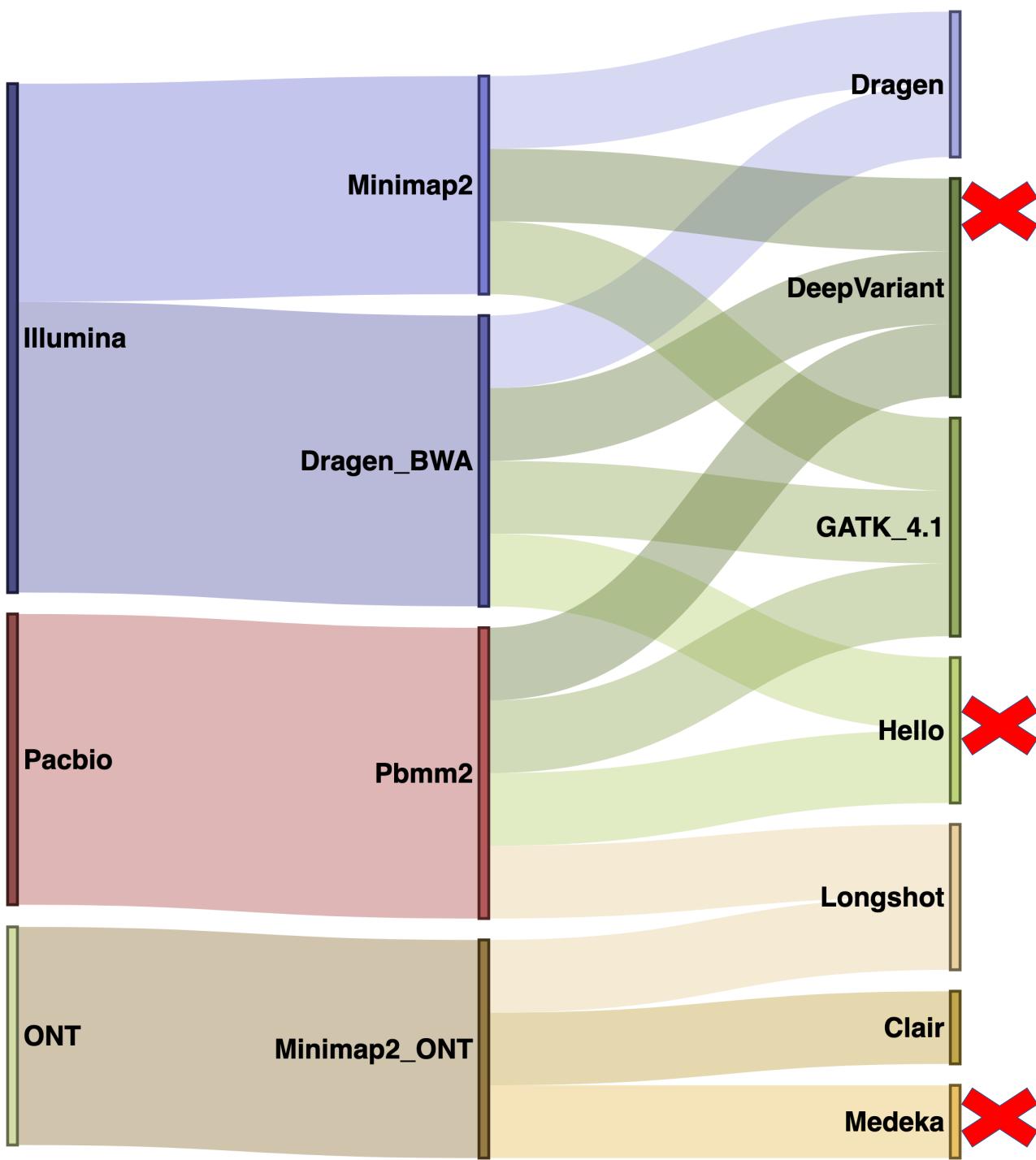
| Tools                     | Version tested |
|---------------------------|----------------|
| Deepvariant               | 3              |
| Clair / Clair3            | 2              |
| Longshot                  | 1              |
| PEPPER-Margin-DeepVariant | 1              |
| Dragen*                   | 4              |
| Octopus                   | 2              |
| GATK                      | 2              |

| Platform | Tools                     |
|----------|---------------------------|
| Pacbio   | Deepvariant               |
| Pacbio   | Clair3                    |
| Pacbio   | Longshot                  |
| Pacbio   | GATK                      |
| ONT      | PEPPER-Margin-DeepVariant |
| ONT      | Clair / Clair3            |
| ONT      | Longshot                  |
| Illumina | Dragen                    |
| Illumina | Deepvariant               |
| Illumina | Octopus                   |
| Illumina | GATK                      |

# Analysis Plan



# Analysis Plan



## Tools used for comparison - Snps

| Platform | Mode                | Tool Version | Recall | Precision | F1_Score |
|----------|---------------------|--------------|--------|-----------|----------|
| Pacbio   | Deepvariant_v1.4.0  | 1.4.0        | 0.9990 | 0.9994    | 0.9992   |
| Pacbio   | Clair3              | 0.1-r9       | 0.9989 | 0.9992    | 0.9991   |
| Pacbio   | Deepvariant_v1.3.0  | 1.3.0        | 0.9986 | 0.9993    | 0.999    |
| Pacbio   | Deepvariant_v0.10.0 | 0.10.0       | 0.9981 | 0.9994    | 0.9987   |
| Pacbio   | Longshot            | 0.4.1        | 0.9829 | 0.9979    | 0.9904   |
| Pacbio   | GATK                | 3.8.1        | 0.9732 | 0.9977    | 0.9853   |

## Tools used for comparison - Snps

| Platform | Mode                | Tool Version | Recall | Precision | F1_Score |
|----------|---------------------|--------------|--------|-----------|----------|
| Pacbio   | Deepvariant_v1.4.0  | 1.4.0        | 0.9990 | 0.9994    | 0.9992   |
| Pacbio   | Clair3              | 0.1-r9       | 0.9989 | 0.9992    | 0.9991   |
| Pacbio   | Deepvariant_v1.3.0  | 1.3.0        | 0.9986 | 0.9993    | 0.999    |
| Pacbio   | Deepvariant_v0.10.0 | 0.10.0       | 0.9981 | 0.9994    | 0.9987   |
| Pacbio   | Longshot            | 0.4.1        | 0.9829 | 0.9979    | 0.9904   |
| Pacbio   | GATK                | 3.8.1        | 0.9732 | 0.9977    | 0.9853   |

| Platform | Mode                      | Tool Version | Recall | Precision | F1_Score |
|----------|---------------------------|--------------|--------|-----------|----------|
| ONT      | Clair3                    | 0.1-r9       | 0.997  | 0.9974    | 0.9972   |
| ONT      | PEPPER-Margin-DeepVariant | 0.8          | 0.9943 | 0.9974    | 0.9959   |
| ONT      | Pepper-Margin-Deepvariant | 0.5          | 0.9976 | 0.9713    | 0.9842   |
| ONT      | Longshot                  | 0.4.1        | 0.9734 | 0.9769    | 0.9752   |
| ONT      | Clair                     | 1            | 0.9911 | 0.9135    | 0.9508   |

# Tools used for comparison - Snps

| Platform | Mode                          | Tool Version | Recall | Precision | F1_Score |
|----------|-------------------------------|--------------|--------|-----------|----------|
| Illumina | Dragen v3.10 ML – Alt Aware   | 3.10         | 0.9974 | 0.9990    | 0.9982   |
| Illumina | Deepvariant_v1.4.0            | 1.4.0        | 0.9965 | 0.9985    | 0.9975   |
| Illumina | Dragen v3.10 – Alt Aware      | 3.10         | 0.9968 | 0.9981    | 0.9975   |
| Illumina | Dragen v3.7 – Alt Aware       | 3.7          | 0.9967 | 0.9975    | 0.9971   |
| Illumina | Deepvariant_v1.3.0            | 1.3.0        | 0.996  | 0.9982    | 0.9971   |
| Illumina | Dragen v3.8 /v3.9 – Alt Aware | 3.8,3.9      | 0.9967 | 0.9975    | 0.9971   |
| Illumina | Octopus_v7.4 RF               | 7.4          | 0.9953 | 0.9988    | 0.997    |
| Illumina | Deepvariant_v0.10.0           | 0.10.0       | 0.9918 | 0.998     | 0.9949   |
| Illumina | Octopus – Alt Bam             | 7.2          | 0.9912 | 0.9984    | 0.9948   |
| Illumina | Dragen                        | 3.6          | 0.9928 | 0.9964    | 0.9946   |
| Illumina | Octopus – Decoy Bam           | 7.2          | 0.9861 | 0.9978    | 0.9919   |
| Illumina | GATK_Relax_noATL              | 3.8.1        | 0.9914 | 0.9912    | 0.9913   |
| Illumina | dragen_Minimap2Bam            | 3.6          | 0.9871 | 0.9948    | 0.9909   |
| Illumina | GATK_Relax                    | 3.8.1        | 0.9913 | 0.9902    | 0.9908   |
| Illumina | GATK_Strict_noATL             | 3.8.1        | 0.9823 | 0.9932    | 0.9877   |
| Illumina | GATK_Strict                   | 3.8.1        | 0.982  | 0.9924    | 0.9871   |
| Illumina | GATK_recal_Pass               | 4.1.9        | 0.9735 | 0.9958    | 0.9845   |
| Illumina | GATK_recal_refinement_noATL   | 3.8.1        | 0.9448 | 0.9983    | 0.9708   |
| Illumina | GATK_recal_refinement         | 3.8.1        | 0.9428 | 0.9981    | 0.9697   |

GATK Indels Strict:

" QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0 "

GATK Relax

" QD < 2.0 || FS > 200.0 "

GATK Pacbio 1base\_indels

"AS\_QD < 5.0"

GATK Pacbio >1base\_indels

"AS\_QD < 2.0"

Dragen Indels

"QUAL < 10.4139 "

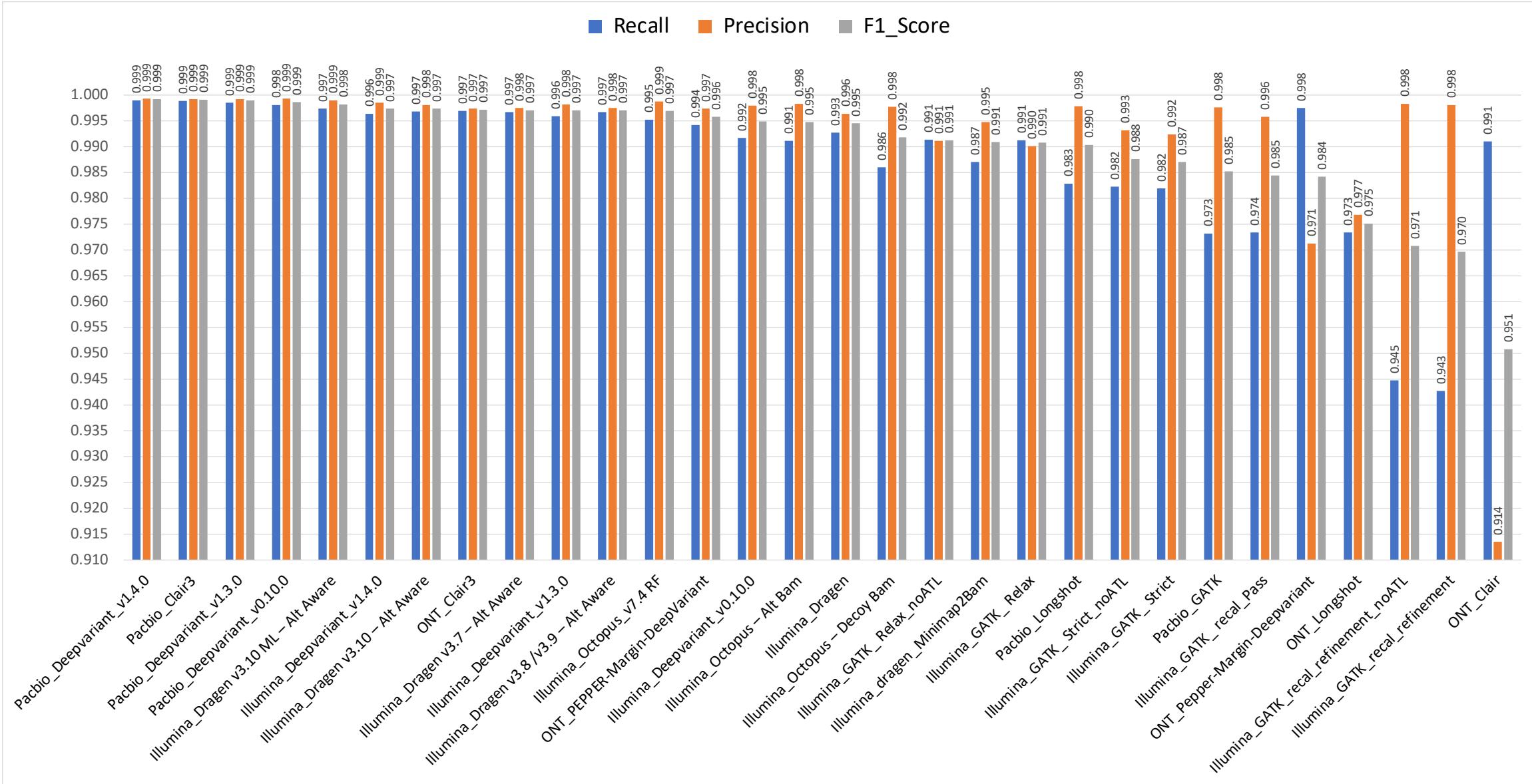
Deepvariant Indels

Default - PASS

# Tools used for comparison - Snps

| Platform | Mode                          | Tool Version | Recall | Precision | F1_Score |
|----------|-------------------------------|--------------|--------|-----------|----------|
| Pacbio   | Deepvariant_v1.4.0            | 1.4.0        | 0.9990 | 0.9994    | 0.9992   |
| Pacbio   | Clair3                        | 0.1-r9       | 0.9989 | 0.9992    | 0.9991   |
| Pacbio   | Deepvariant_v1.3.0            | 1.3.0        | 0.9986 | 0.9993    | 0.999    |
| Pacbio   | Deepvariant_v0.10.0           | 0.10.0       | 0.9981 | 0.9994    | 0.9987   |
| Illumina | Dragen v3.10 ML – Alt Aware   | 3.10         | 0.9974 | 0.9990    | 0.9982   |
| Illumina | Deepvariant_v1.4.0            | 1.4.0        | 0.9965 | 0.9985    | 0.9975   |
| Illumina | Dragen v3.10 – Alt Aware      | 3.10         | 0.9968 | 0.9981    | 0.9975   |
| ONT      | Clair3                        | 0.1-r9       | 0.997  | 0.9974    | 0.9972   |
| Illumina | Dragen v3.7 – Alt Aware       | 3.7          | 0.9967 | 0.9975    | 0.9971   |
| Illumina | Deepvariant_v1.3.0            | 1.3.0        | 0.996  | 0.9982    | 0.9971   |
| Illumina | Dragen v3.8 /v3.9 – Alt Aware | 3.8,3.9      | 0.9967 | 0.9975    | 0.9971   |
| Illumina | Octopus_v7.4 RF               | 7.4          | 0.9953 | 0.9988    | 0.997    |
| ONT      | PEPPER-Margin-DeepVariant     | 0.8          | 0.9943 | 0.9974    | 0.9959   |
| Illumina | Deepvariant_v0.10.0           | 0.10.0       | 0.9918 | 0.998     | 0.9949   |
| Illumina | Octopus – Alt Bam             | 7.2          | 0.9912 | 0.9984    | 0.9948   |
| Illumina | Dragen                        | 3.6          | 0.9928 | 0.9964    | 0.9946   |
| Illumina | Octopus – Decoy Bam           | 7.2          | 0.9861 | 0.9978    | 0.9919   |
| Illumina | GATK_Relax_noATL              | 3.8.1        | 0.9914 | 0.9912    | 0.9913   |
| Illumina | dragen_Minimap2Bam            | 3.6          | 0.9871 | 0.9948    | 0.9909   |
| Illumina | GATK_Relax                    | 3.8.1        | 0.9913 | 0.9902    | 0.9908   |
| Pacbio   | Longshot                      | 0.4.1        | 0.9829 | 0.9979    | 0.9904   |
| Illumina | GATK_Strict_noATL             | 3.8.1        | 0.9823 | 0.9932    | 0.9877   |
| Illumina | GATK_Strict                   | 3.8.1        | 0.982  | 0.9924    | 0.9871   |
| Pacbio   | GATK                          | 3.8.1        | 0.9732 | 0.9977    | 0.9853   |
| Illumina | GATK_recal_Pass               | 4.1.9        | 0.9735 | 0.9958    | 0.9845   |
| ONT      | Pepper-Margin-Deepvariant     | 0.5          | 0.9976 | 0.9713    | 0.9842   |
| ONT      | Longshot                      | 0.4.1        | 0.9734 | 0.9769    | 0.9752   |
| Illumina | GATK_recal_refinement_noATL   | 3.8.1        | 0.9448 | 0.9983    | 0.9708   |
| Illumina | GATK_recal_refinement         | 3.8.1        | 0.9428 | 0.9981    | 0.9697   |
| ONT      | Clair                         | 1            | 0.9911 | 0.9135    | 0.9508   |

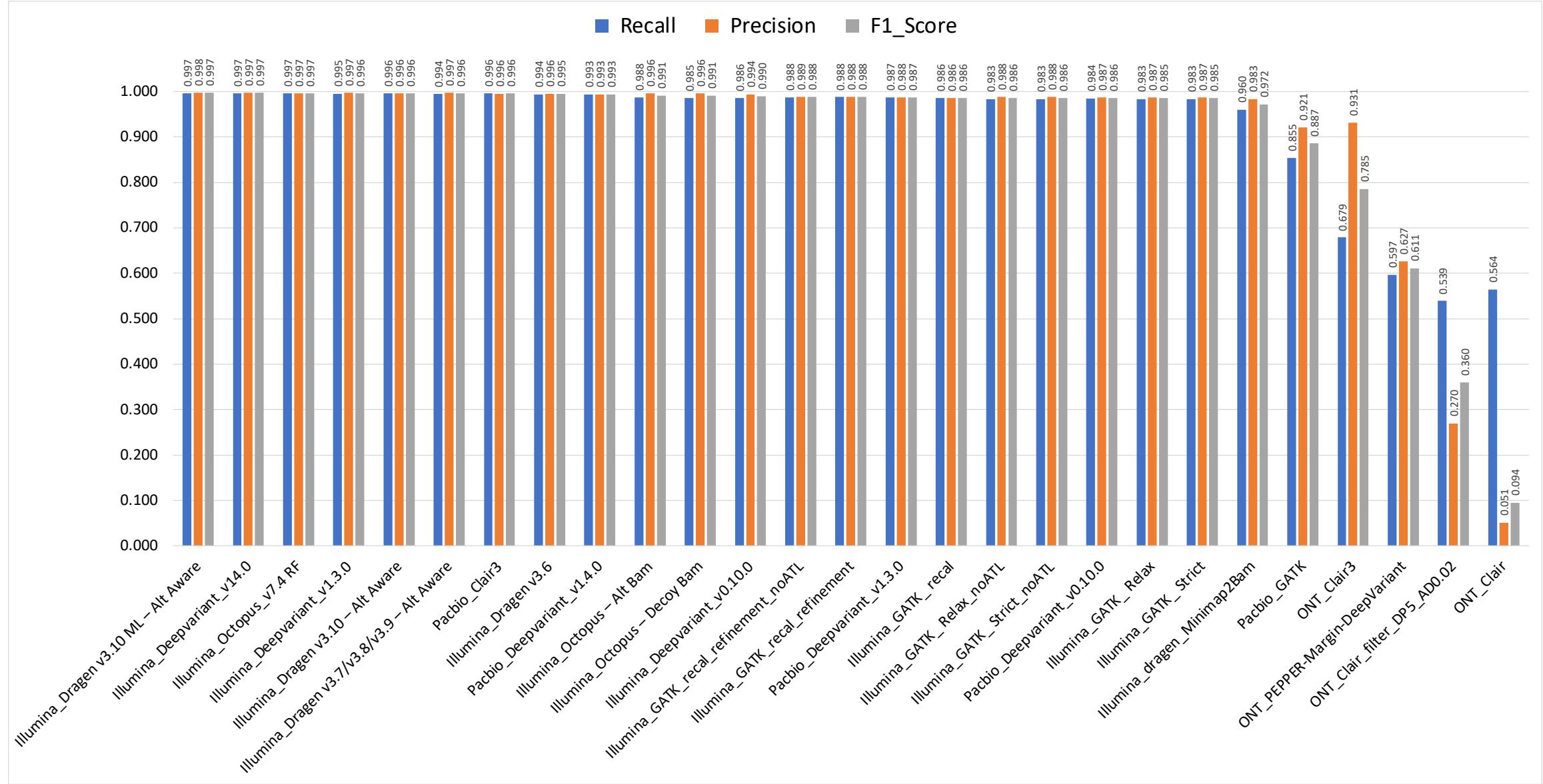
# Tools used for comparison - Snps



# Tools used for comparison - Indels

| <i>Platform</i> | <i>Mode</i>                       | <i>Tool Version</i> | <i>Recall</i> | <i>Precision</i> | <i>F1_Score</i> |
|-----------------|-----------------------------------|---------------------|---------------|------------------|-----------------|
| Illumina        | Dragen v3.10 ML – Alt Aware       | 3.10                | 0.9969        | 0.9981           | 0.9975          |
| Illumina        | Deepvariant_v14.0                 | 1.4.0               | 0.9967        | 0.9973           | 0.9970          |
| Illumina        | Octopus_v7.4 RF                   | 7.4                 | 0.9966        | 0.9968           | 0.9967          |
| Illumina        | Deepvariant_v1.3.0                | 1.3.0               | 0.9949        | 0.9974           | 0.9962          |
| Illumina        | Dragen v3.10 – Alt Aware          | 3.10                | 0.9957        | 0.9962           | 0.996           |
| Illumina        | Dragen v3.7/v3.8/v3.9 – Alt Aware | 3.7, 3.8, 3.9       | 0.9944        | 0.9974           | 0.9959          |
| Pacbio          | Clair3                            | 0.1-r9              | 0.996         | 0.9956           | 0.9958          |
| Illumina        | Dragen v3.6                       | 3.6                 | 0.9936        | 0.9955           | 0.9945          |
| Pacbio          | Deepvariant_v1.4.0                | 1.4.0               | 0.9935        | 0.9931           | 0.9933          |
| Illumina        | Octopus – Alt Bam                 | 7.2                 | 0.9876        | 0.9962           | 0.9912          |
| Illumina        | Octopus – Decoy Bam               | 7.2                 | 0.9854        | 0.996            | 0.9906          |
| Illumina        | Deepvariant_v0.10.0               | 0.10.0              | 0.9864        | 0.9943           | 0.9904          |
| Illumina        | GATK_recal_refinement_noATL       | 3.8.1               | 0.9878        | 0.989            | 0.9884          |
| Illumina        | GATK_recal_refinement             | 3.8.1               | 0.9881        | 0.988            | 0.9881          |
| Pacbio          | Deepvariant_v1.3.0                | 1.3.0               | 0.9869        | 0.9875           | 0.9872          |
| Illumina        | GATK_recal                        | 4.1.9               | 0.9864        | 0.9858           | 0.9861          |
| Illumina        | GATK_Relax_noATL                  | 3.8.1               | 0.9833        | 0.9879           | 0.9856          |
| Illumina        | GATK_Strict_noATL                 | 3.8.1               | 0.9833        | 0.9879           | 0.9856          |
| Pacbio          | Deepvariant_v0.10.0               | 0.10.1              | 0.9843        | 0.9869           | 0.9856          |
| Illumina        | GATK_Relax                        | 3.8.1               | 0.9832        | 0.9874           | 0.9853          |
| Illumina        | GATK_Strict                       | 3.8.1               | 0.9832        | 0.9874           | 0.9853          |
| Illumina        | dragen_Minimap2Bam                | 3.6                 | 0.9604        | 0.9834           | 0.9718          |
| Pacbio          | GATK                              | 3.8.1               | 0.8546        | 0.9214           | 0.8868          |
| ONT             | Clair3                            | 0.1-r9              | 0.6788        | 0.931            | 0.7851          |
| ONT             | PEPPER-Margin-DeepVariant         | 0.8                 | 0.5970        | 0.6268           | 0.6114          |
| ONT             | Clair_filter_DP5_AD0.02           | 1                   | 0.5393        | 0.2698           | 0.3597          |
| ONT             | Clair                             | 1                   | 0.564         | 0.0513           | 0.0941          |

# Tools used for comparison - Indels



## Tools used for comparison - Mappers

| Platform | Mode               | Tool Version | Mapper | Recall | Precision | F1_Score |
|----------|--------------------|--------------|--------|--------|-----------|----------|
| Illumina | Octopus_v7.4 RF    | 7.4          | Dragen | 0.9953 | 0.9988    | 0.997    |
| Illumina | Octopus_v7.4 RF    | 7.4          | BWA    | 0.9920 | 0.9980    | 0.9950   |
| Illumina | Deepvariant_v1.3.0 | 1.3.0        | Dragen | 0.996  | 0.9982    | 0.9971   |
| Illumina | Deepvariant_v1.3.0 | 1.3.0        | BWA    | 0.993  | 0.9984    | 0.9957   |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | Dragen | 0.9965 | 0.9985    | 0.9975   |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | BWA    | 0.9937 | 0.9989    | 0.9963   |

## Tools used for comparison - Mappers

| Platform | Mode               | Tool Version | Mapper | Recall | Precision | F1_Score |
|----------|--------------------|--------------|--------|--------|-----------|----------|
| Illumina | Octopus_v7.4 RF    | 7.4          | Dragen | 0.9953 | 0.9988    | 0.997    |
| Illumina | Octopus_v7.4 RF    | 7.4          | BWA    | 0.9920 | 0.9980    | 0.9950   |
| Illumina | Deepvariant_v1.3.0 | 1.3.0        | Dragen | 0.996  | 0.9982    | 0.9971   |
| Illumina | Deepvariant_v1.3.0 | 1.3.0        | BWA    | 0.993  | 0.9984    | 0.9957   |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | Dragen | 0.9965 | 0.9985    | 0.9975   |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | BWA    | 0.9937 | 0.9989    | 0.9963   |

Snetieon

# Sentieon - DNAScope



bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

[Follow this preprint](#)

## **DNAScope: High accuracy small variant calling using machine learning**

Donald Freed, Renke Pan, Haodong Chen, Zhipan Li, Jinnan Hu, Rafael Aldana

**doi:** <https://doi.org/10.1101/2022.05.20.492556>

This article is a preprint and has not been certified by peer review [what does this mean?].



[Abstract](#)

[Full Text](#)

[Info/History](#)

[Metrics](#)

[Preview PDF](#)

# Sentieon - DNAScope



bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

[Follow this preprint](#)

## DNAscope: High accuracy small variant calling using machine learning

Donald Freed, Renke Pan, Haodong Chen, Zhipan Li, Jinnan Hu, Rafael Aldana

**doi:** <https://doi.org/10.1101/2022.05.20.492556>

This article is a preprint and has not been certified by peer review [what does this mean?].



[Abstract](#)

[Full Text](#)

[Info/History](#)

[Metrics](#)

[Preview PDF](#)

DNAscope achieves an F1-score of 99.57% for SNPs and 99.46% for INDELS

# Sentieon – DNAScope Long Read



bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

[Follow this preprint](#)

## **Sentieon DNAScope LongRead – A highly Accurate, Fast, and Efficient Pipeline for Germline Variant Calling from PacBio HiFi reads**

Donald Freed, William J. Rowell, Aaron M. Wenger, Zhipan Li

**doi:** <https://doi.org/10.1101/2022.06.01.494452>

This article is a preprint and has not been certified by peer review [what does this mean?].



[Abstract](#)

[Full Text](#)

[Info/History](#)

[Metrics](#)

[Preview PDF](#)

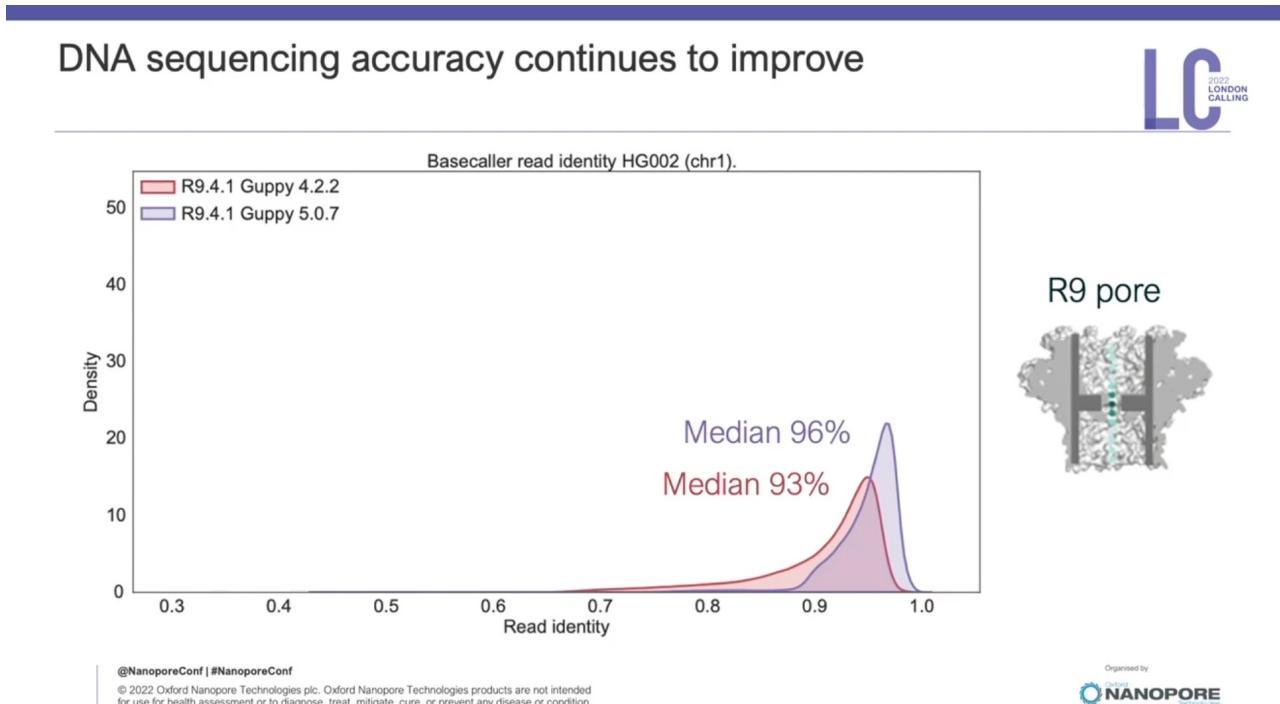
The mean F1-score for the three pFDA samples is 0.9988 for the DNAScope LongRead pipeline compared to 0.9986 for the pFDA winning pipeline. For HG002, F1 score of 0.9989

# Why Long-read data for DNaseq?

- Pacbio or Oxford Nanopore:
  - SNVs
  - SVs
  - Base Mod - 5mC and/or 6mA
- Pacbio: Hifi : Most Accurate Long reads
  - > Q40 Read quality
    - Most accurate variant calls

# Why Long-read data for DNaseq?

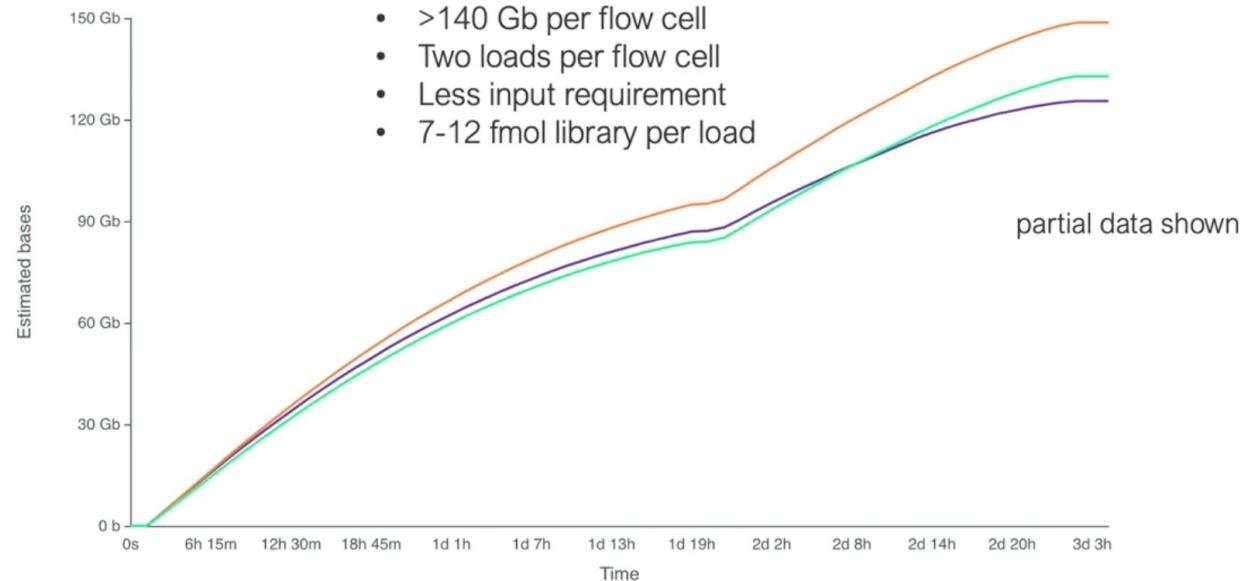
- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!



# Why Long-read data for DNaseq?

- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!

Kit14 high-speed (400 bps) runs are high throughput



@NanoporeConf | #NanoporeConf

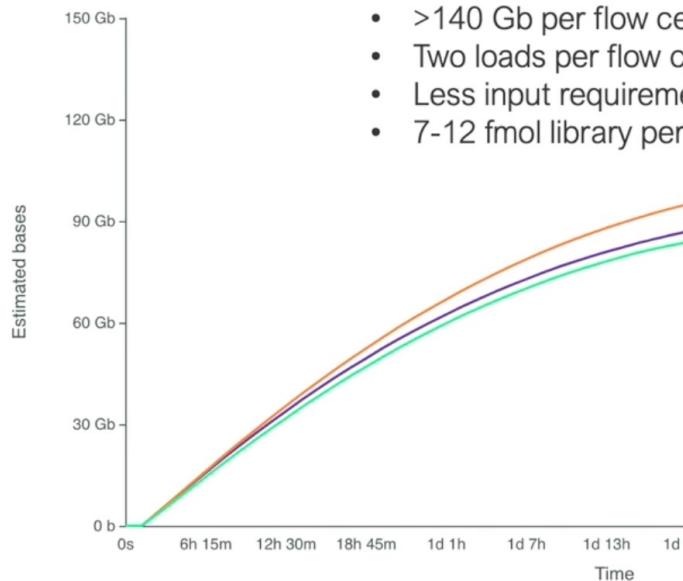
© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Organised by  
OXFORD NANOPORE TECHNOLOGIES

# Why Long-read data for DNaseq?

- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!

Kit14 high-speed (400 bps) runs are high throughput

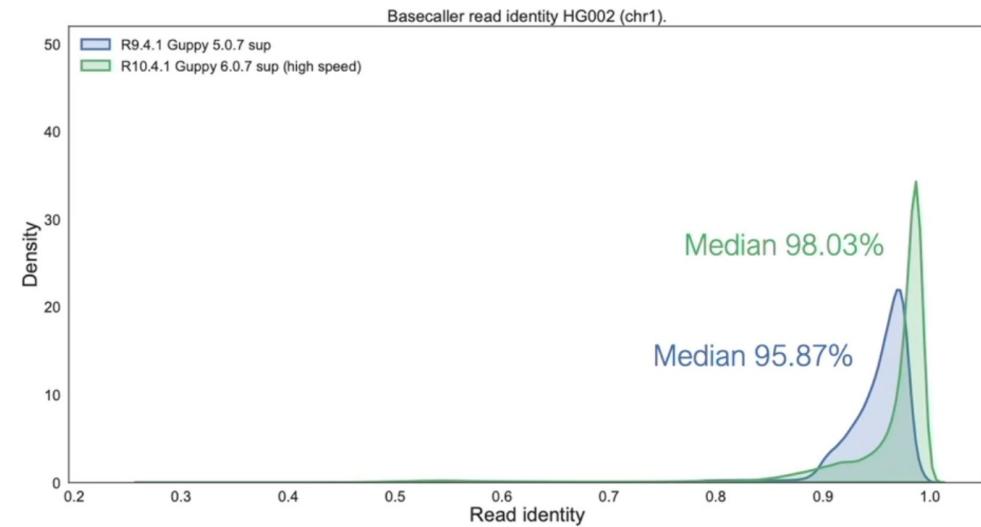


@NanoporeConf | #NanoporeConf  
© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Miten Jain: London Calling 2022



Kit14 high-speed (400 bps) runs have high alignment identity



© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.



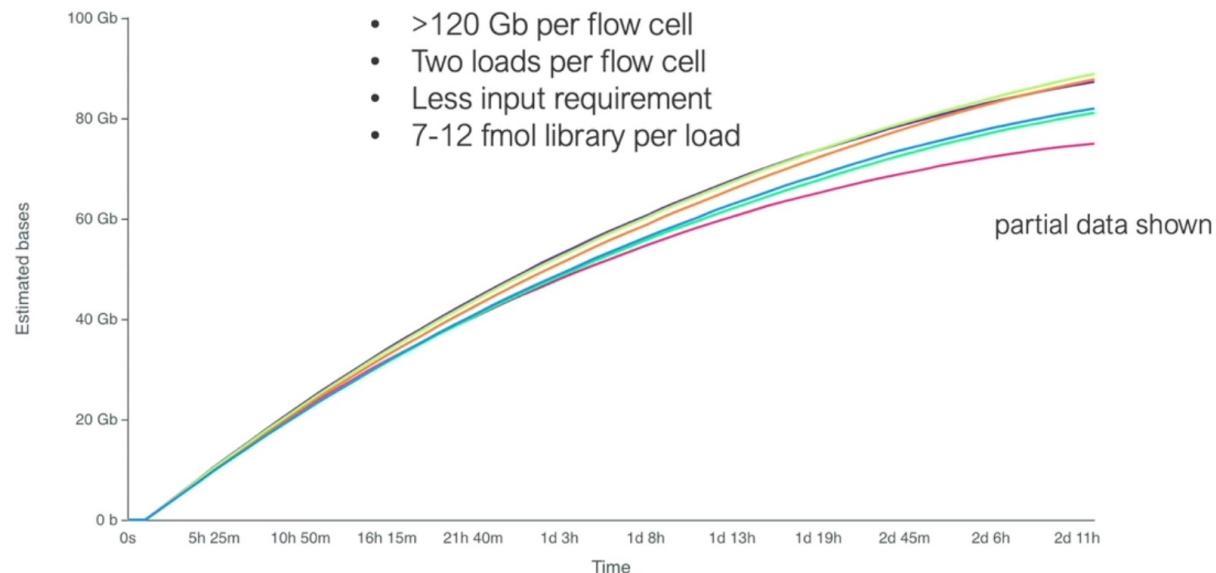
Oxford NANOPORE Technologies

# Why Long-read data for DNaseq?

- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!

---

Kit14 low-speed (260 bps) runs yield good throughput



@NanoporeConf | #NanoporeConf

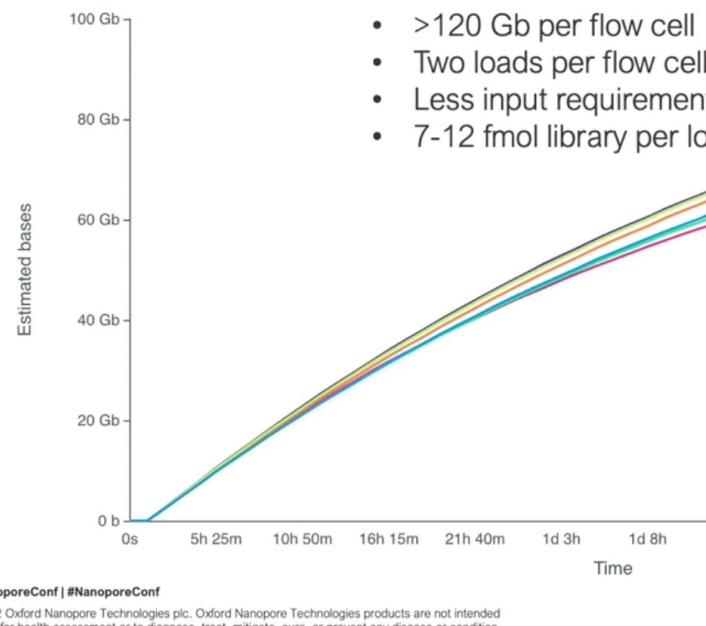
© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Organised by  
 NANOPORE  
Technologies

# Why Long-read data for DNaseq?

- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!

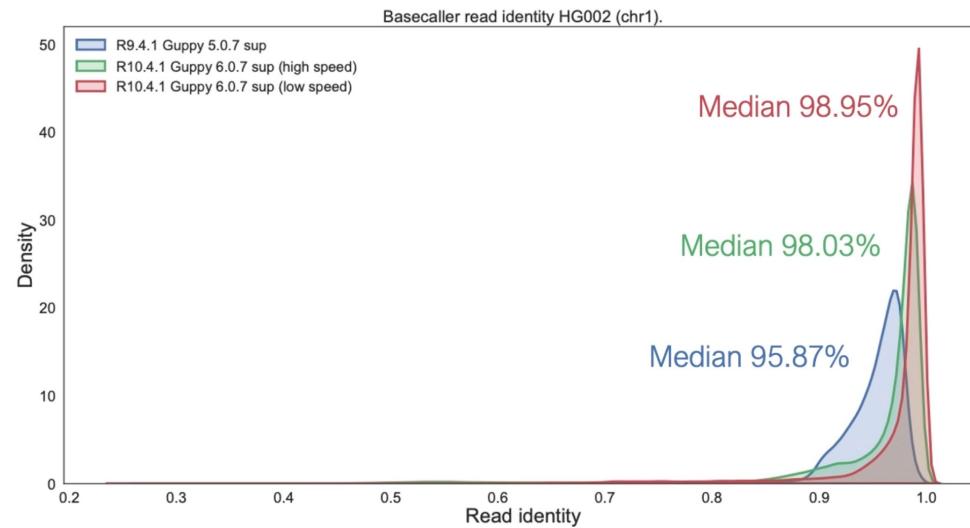
Kit14 low-speed (260 bps) runs yield good throughput



Miten Jain: London Calling 2022



Kit14 low-speed (260 bps) runs yield higher alignment identity



© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.



# Why Long-read data for DNaseq?

- Oxford Nanopore: R10.4.1, Kit 14, and duplex data
  - Ultra long reads
  - Quality can increase after sequencing!

Good SNV calling performance – improvements forthcoming



| Sample                                       | Type  | Recall   | Precision | F1-Score |
|--|-------|----------|-----------|----------|
| HG002 chr20<br>R10.4.1 + Kit14<br>High-speed | INDEL | 0.829424 | 0.911302  | 0.868437 |
|  | SNP   | 0.998822 | 0.998781  | 0.998802 |
| HG002 chr20<br>R10.4.1 + Kit14<br>Low-speed  | INDEL | 0.852345 | 0.920583  | 0.885151 |
|  | SNP   | 0.999005 | 0.999411  | 0.999208 |

nature|methods

@NanoporeConf | #NanoporeConf

© 2022 Oxford Nanopore Technologies plc. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Haplotype-aware variant calling with  
PEPPER-Margin-DeepVariant enables high  
accuracy in nanopore long-reads

\* PEPPER-Margin-DeepVariant training underway

Google Health UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

Organised by  
Oxford NANOPORE Technologies

# CHM13

## T2T-CHM13v2.0

**Description:** T2T CHM13v2.0 Telomere-to-Telomere assembly of the CHM13 cell line, with chrY from NA24385

**Organism name:** [Homo sapiens \(human\)](#)

**BioProject:** [PRJNA559484](#)

**Submitter:** T2T Consortium

**Date:** 2022/01/24

**Assembly level:** Complete Genome

**Genome representation:** full

**GenBank assembly accession:** GCA\_009914755.4 (latest)

**RefSeq assembly accession:** GCF\_009914755.1 (latest)

**RefSeq assembly and GenBank assembly identical:** no ([hide details](#))

- Only in GenBank: chromosome MT (in non-nuclear assembly-unit)
- Data displayed for RefSeq version

**Expected final version:** no

**Genome coverage:** 30x

IDs: 11828891 [UID] 31127148 [GenBank] 31865168 [RefSeq]

**History** ([Show revision history](#))

### Comment

#### ▼ [Genome-Annotation-Data](#)

```
##Genome-Annotation-Data-START##  
Annotation Provider::NCBI  
Annotation Status::Full annotation  
Annotation Name::Homo sapiens Annotation Release 110  
Annotation Version::110  
Annotation Pipeline::NCBI eukaryotic genome annotation pipeline  
Annotation Software Version::9.0  
Annotation Method::Best-placed RefSeq; Gnomon  
Features Annotated::Gene; mRNA; CDS; ncRNA  
##Genome-Annotation-Data-END##
```

### Global statistics

|  |               |
|--|---------------|
| Total sequence length                    | 3,117,275,501 |
| Total ungapped length                    | 3,117,275,501 |
| Total number of chromosomes and plasmids | 24            |

# CHM13

## T2T-CHM13v2.0

**Description:** T2T CHM13v2.0 Telomere-to-Telomere assembly of the CHM13 cell line, with chrY from NA24385

**Organism name:** [Homo sapiens \(human\)](#)

**BioProject:** [PRJNA559484](#)

**Submitter:** T2T Consortium

**Date:** 2022/01/24

**Assembly level:** Complete Genome

**Genome representation:** full

**GenBank assembly accession:** GCA\_009914755.4 (latest)

**RefSeq assembly accession:** GCF\_009914755.1 (latest)

**RefSeq assembly and GenBank assembly identical:** no ([hide details](#))

- Only in GenBank: chromosome MT (in non-nuclear assembly-unit)
- Data displayed for RefSeq version

**Expected final version:** no

**Genome coverage:** 30x

IDs: 11828891 [UID] 31127148 [GenBank] 31865168 [RefSeq]

**History** ([Show revision history](#))

### Comment

#### ▼ [Genome-Annotation-Data](#)

```
##Genome-Annotation-Data-START##  
Annotation Provider::NCBI  
Annotation Status::Full annotation  
Annotation Name::Homo sapiens Annotation Release 110  
Annotation Version::110  
Annotation Pipeline::NCBI eukaryotic genome annotation pipeline  
Annotation Software Version::9.0  
Annotation Method::Best-placed RefSeq; Gnomon  
Features Annotated::Gene; mRNA; CDS; ncRNA  
##Genome-Annotation-Data-END##
```

### Global statistics

|  |               |
|--|---------------|
| Total sequence length                    | 3,117,275,501 |
| Total ungapped length                    | 3,117,275,501 |
| Total number of chromosomes and plasmids | 24            |

## CHM13

| <b><i>Chromosome</i></b> | <b><i>CHM13v2</i></b> | <b><i>hg38</i></b> | <b><i>Difference in bp</i></b> |
|--------------------------|-----------------------|--------------------|--------------------------------|
| chr1                     | 248,387,328           | 248,956,422        | (569,094)                      |
| chr2                     | 242,696,752           | 242,193,529        | 503,223                        |
| chr3                     | 201,105,948           | 198,295,559        | 2,810,389                      |
| chr4                     | 193,574,945           | 190,214,555        | 3,360,390                      |
| chr5                     | 182,045,439           | 181,538,259        | 507,180                        |
| chr6                     | 172,126,628           | 170,805,979        | 1,320,649                      |
| chr7                     | 160,567,428           | 159,345,973        | 1,221,455                      |
| chr8                     | 146,259,331           | 145,138,636        | 1,120,695                      |
| chr9                     | 150,617,247           | 138,394,717        | 12,222,530                     |
| chr10                    | 134,758,134           | 133,797,422        | 960,712                        |
| chr11                    | 135,127,769           | 135,086,622        | 41,147                         |
| chr12                    | 133,324,548           | 133,275,309        | 49,239                         |
| chr13                    | 113,566,686           | 114,364,328        | (797,642)                      |
| chr14                    | 101,161,492           | 107,043,718        | (5,882,226)                    |
| chr15                    | 99,753,195            | 101,991,189        | (2,237,994)                    |
| chr16                    | 96,330,374            | 90,338,345         | 5,992,029                      |
| chr17                    | 84,276,897            | 83,257,441         | 1,019,456                      |
| chr18                    | 80,542,538            | 80,373,285         | 169,253                        |
| chr19                    | 61,707,364            | 58,617,616         | 3,089,748                      |
| chr20                    | 66,210,255            | 64,444,167         | 1,766,088                      |
| chr21                    | 45,090,682            | 46,709,983         | (1,619,301)                    |
| chr22                    | 51,324,926            | 50,818,468         | 506,458                        |
| chrX                     | 154,259,566           | 156,040,895        | (1,781,329)                    |
| chrY                     | 62,460,029            | 57,227,415         | 5,232,614                      |
| chrM                     | 16,569                | 16,569             | -                              |



bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Follow this preprint

## Improved sequence mapping using a complete reference genome and lift-over

Nae-Chyun Chen, Luis F Paulin, Fritz J Sedlazeck, Sergey Koren, Adam M Phillippy, Ben Langmead

**doi:** <https://doi.org/10.1101/2022.04.27.489683>

This article is a preprint and has not been certified by peer review [what does this mean?].



Abstract

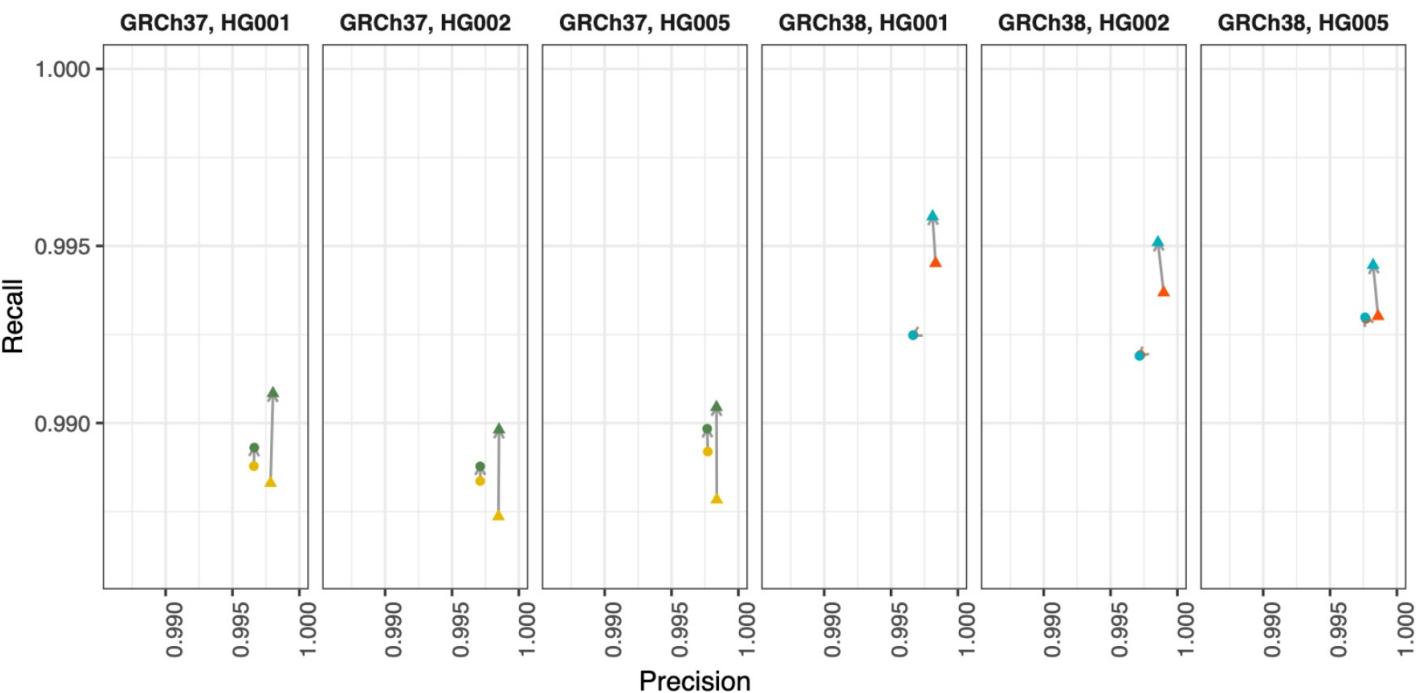
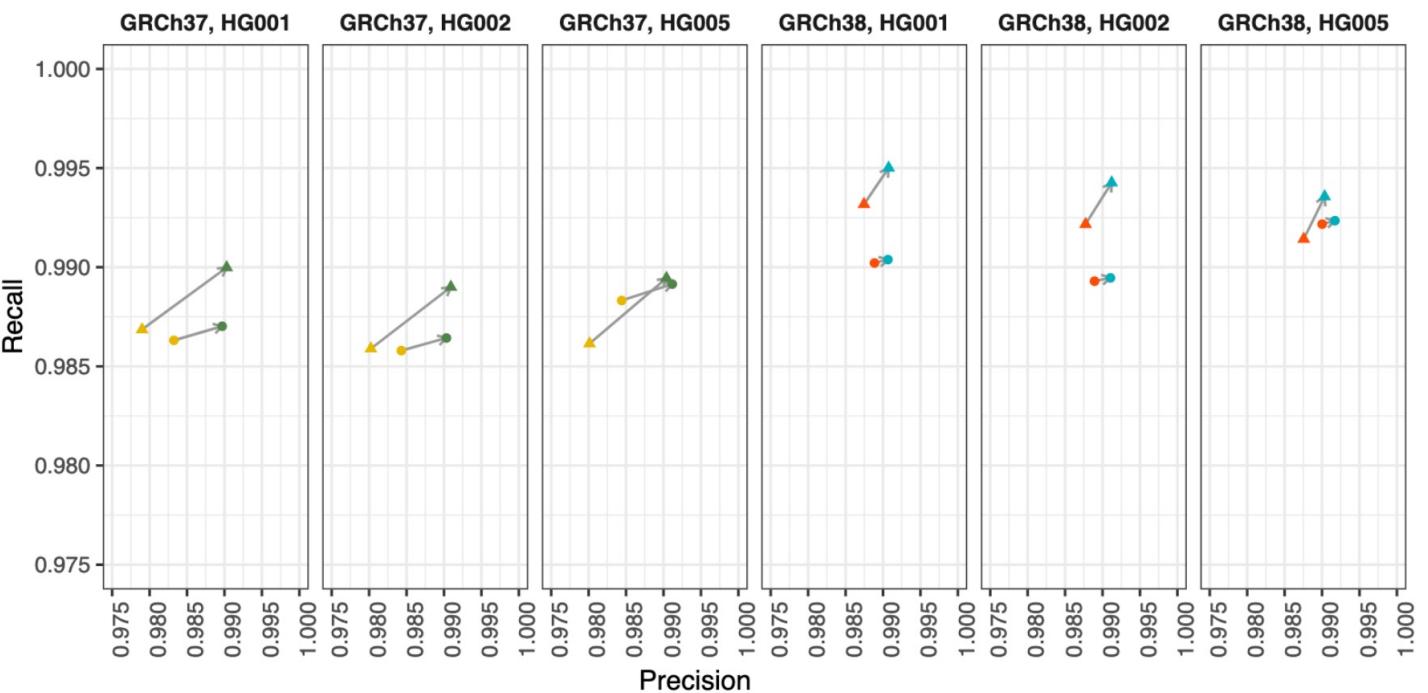
Full Text

Info/History

Metrics

Preview PDF

CHM13



CHM1

**Table S2:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM-GATK-HaplotypeCaller in all GIAB v4.2.1 regions<sup>22</sup>

| Sample | Method          | Type  | Recall | Precision | $F_1$  | TP      | FN    | FP    |
|--------|-----------------|-------|--------|-----------|--------|---------|-------|-------|
| HG002  | GRCh38          | SNP   | 0.9922 | 0.9877    | 0.9899 | 3338776 | 26350 | 41484 |
| HG002  | CHM13-to-GRCh38 | SNP   | 0.9943 | 0.9912    | 0.9927 | 3345806 | 19320 | 29656 |
| HG002  | GRCh37          | SNP   | 0.9859 | 0.9802    | 0.9831 | 3305380 | 47305 | 66646 |
| HG002  | CHM13-to-GRCh37 | SNP   | 0.9890 | 0.9909    | 0.9900 | 3315802 | 36883 | 30406 |
| HG002  | GRCh38          | INDEL | 0.9893 | 0.9889    | 0.9891 | 519842  | 5625  | 6059  |
| HG002  | CHM13-to-GRCh38 | INDEL | 0.9895 | 0.9911    | 0.9903 | 519929  | 5538  | 4887  |
| HG002  | GRCh37          | INDEL | 0.9858 | 0.9843    | 0.9851 | 514968  | 7421  | 8551  |
| HG002  | CHM13-to-GRCh37 | INDEL | 0.9864 | 0.9903    | 0.9884 | 515300  | 7089  | 5244  |

**Table S3:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM-DeepVariant in all GIAB v4.2.1 regions<sup>22</sup>

| Sample | Method          | Type  | Recall | Precision | $F_1$  | TP      | FN    | FP   |
|--------|-----------------|-------|--------|-----------|--------|---------|-------|------|
| HG002  | GRCh38          | SNP   | 0.9937 | 0.9990    | 0.9963 | 3343863 | 21263 | 3436 |
| HG002  | CHM13-to-GRCh38 | SNP   | 0.9951 | 0.9985    | 0.9968 | 3348636 | 16490 | 4881 |
| HG002  | GRCh37          | SNP   | 0.9874 | 0.9985    | 0.9929 | 3310312 | 42373 | 5059 |
| HG002  | CHM13-to-GRCh37 | SNP   | 0.9898 | 0.9985    | 0.9941 | 3318525 | 34160 | 4935 |
| HG002  | GRCh38          | INDEL | 0.9919 | 0.9972    | 0.9946 | 521223  | 4244  | 1506 |
| HG002  | CHM13-to-GRCh38 | INDEL | 0.9919 | 0.9972    | 0.9945 | 521209  | 4258  | 1548 |
| HG002  | GRCh37          | INDEL | 0.9884 | 0.9971    | 0.9927 | 516310  | 6079  | 1560 |
| HG002  | CHM13-to-GRCh37 | INDEL | 0.9888 | 0.9971    | 0.9929 | 516527  | 5862  | 1561 |

CHM13

**Table S2:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM–GATK-HaplotypeCaller in all GIAB v4.2.1 regions<sup>22</sup>

| Sample | Method          | Type  | Recall | Precision | $F_1$  | TP      | FN    | FP    |
|--------|-----------------|-------|--------|-----------|--------|---------|-------|-------|
| HG002  | GRCh38          | SNP   | 0.9922 | 0.9877    | 0.9899 | 3338776 | 26350 | 41484 |
| HG002  | CHM13-to-GRCh38 | SNP   | 0.9943 | 0.9912    | 0.9927 | 3345806 | 19320 | 29656 |
| HG002  | GRCh37          | SNP   | 0.9859 | 0.9802    | 0.9831 | 3305380 | 47305 | 66646 |
| HG002  | CHM13-to-GRCh37 | SNP   | 0.9890 | 0.9909    | 0.9900 | 3315802 | 36883 | 30406 |
| HG002  | GRCh38          | INDEL | 0.9893 | 0.9889    | 0.9891 | 519842  | 5625  | 6059  |
| HG002  | CHM13-to-GRCh38 | INDEL | 0.9895 | 0.9911    | 0.9903 | 519929  | 5538  | 4887  |
| HG002  | GRCh37          | INDEL | 0.9858 | 0.9843    | 0.9851 | 514968  | 7421  | 8551  |
| HG002  | CHM13-to-GRCh37 | INDEL | 0.9864 | 0.9903    | 0.9884 | 515300  | 7089  | 5244  |

**Table S3:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM–DeepVariant in all GIAB v4.2.1 regions<sup>22</sup>

| Sample | Method          | Type  | Recall | Precision | $F_1$  | TP      | FN    | FP   |
|--------|-----------------|-------|--------|-----------|--------|---------|-------|------|
| HG002  | GRCh38          | SNP   | 0.9937 | 0.9990    | 0.9963 | 3343863 | 21263 | 3436 |
| HG002  | CHM13-to-GRCh38 | SNP   | 0.9951 | 0.9985    | 0.9968 | 3348636 | 16490 | 4881 |
| HG002  | GRCh37          | SNP   | 0.9874 | 0.9985    | 0.9929 | 3310312 | 42373 | 5059 |
| HG002  | CHM13-to-GRCh37 | SNP   | 0.9898 | 0.9985    | 0.9941 | 3318525 | 34160 | 4935 |
| HG002  | GRCh38          | INDEL | 0.9919 | 0.9972    | 0.9946 | 521223  | 4244  | 1506 |
| HG002  | CHM13-to-GRCh38 | INDEL | 0.9919 | 0.9972    | 0.9945 | 521209  | 4258  | 1548 |
| HG002  | GRCh37          | INDEL | 0.9884 | 0.9971    | 0.9927 | 516310  | 6079  | 1560 |
| HG002  | CHM13-to-GRCh37 | INDEL | 0.9888 | 0.9971    | 0.9929 | 516527  | 5862  | 1561 |

CHM13

**Table S2:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM-GATK-HaplotypeCaller in all GIAB v4.2.1 regions<sup>22</sup>

| Sample | Method          | Type  | Recall | Precision | $F_1$  | TP      | FN    | FP    |
|--------|-----------------|-------|--------|-----------|--------|---------|-------|-------|
| HG002  | GRCh38          | SNP   | 0.9922 | 0.9877    | 0.9899 | 3338776 | 26350 | 41484 |
| HG002  | CHM13-to-GRCh38 | SNP   | 0.9943 | 0.9912    | 0.9927 | 3345806 | 19320 | 29656 |
| HG002  | GRCh37          | SNP   | 0.9859 | 0.9802    | 0.9831 | 3305380 | 47305 | 66646 |
| HG002  | CHM13-to-GRCh37 | SNP   | 0.9890 | 0.9909    | 0.9900 | 3315802 | 36883 | 30406 |
| HG002  | GRCh38          | INDEL | 0.9893 | 0.9889    | 0.9891 | 519842  | 5625  | 6059  |
| HG002  | CHM13-to-GRCh38 | INDEL | 0.9895 | 0.9911    | 0.9903 | 519929  | 5538  | 4887  |
| HG002  | GRCh37          | INDEL | 0.9858 | 0.9843    | 0.9851 | 514968  | 7421  | 8551  |
| HG002  | CHM13-to-GRCh37 | INDEL | 0.9864 | 0.9903    | 0.9884 | 515300  | 7089  | 5244  |

**Table S3:** Small variant calling accuracy for 30× WGS datasets using BWA-MEM-DeepVariant in all GIAB v4.2.1 regions<sup>22</sup>

| Sample   | Method             | Type         | Recall | Precision | $F_1$     | TP       | FN    | FP   |
|----------|--------------------|--------------|--------|-----------|-----------|----------|-------|------|
| HG002    | GRCh38             | SNP          | 0.9937 | 0.9990    | 0.9963    | 3343863  | 21263 | 3436 |
| HG002    | CHM13-to-GRCh38    | SNP          | 0.9951 | 0.9985    | 0.9968    | 3348636  | 16490 | 4881 |
| <hr/>    |                    |              |        |           |           |          |       |      |
| Platform | Mode               | Tool Version | Mapper | Recall    | Precision | F1_Score |       |      |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | Dragen | 0.9965    | 0.9985    | 0.9975   |       |      |
| Illumina | Deepvariant_v1.4.0 | 1.4.0        | BWA    | 0.9937    | 0.9989    | 0.9963   |       |      |

# *Somatic Analysis*

# SEQC2 Consortium

nature biotechnology

Explore content ▾ About the journal ▾ Publish with us ▾

---

[nature](#) > [nature biotechnology](#) > [resources](#) > [article](#)

Resource | [Published: 09 September 2021](#)

## Establishing community reference samples, data and call sets for benchmarking cancer mutation detection using whole-genome sequencing

[Li Tai Fang](#), [Bin Zhu](#), ... [The Somatic Mutation Working Group of Sequencing Quality Control Phase II](#)

[Consortium](#) + Show authors

[Nature Biotechnology](#) 39, 1151–1160 (2021) | [Cite this article](#)

7155 Accesses | 7 Citations | 103 Altmetric | [Metrics](#)

# SEQC2 Consortium

nature biotechnology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature biotechnology](#) > [resources](#) > [article](#)

Resource | Published: 09 September 2021

## Establishing community reference sample call sets for benchmarking cancer mutation detection using whole-genome sequencing

[Li Tai Fang](#), [Bin Zhu](#), ... [The Somatic Mutation Working Group of Sequencing Consortium](#)

[Consortium](#) + Show authors

[Nature Biotechnology](#) 39, 1151–1160 (2021) | [Cite this article](#)

7155 Accesses | 7 Citations | 103 Altmetric | [Metrics](#)

nature biotechnology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature biotechnology](#) > [analyses](#) > [article](#)

Analysis | Published: 09 September 2021

## Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing

[Wenming Xiao](#)✉, [Luyao Ren](#), ... [Leming Shi](#)✉ + Show authors

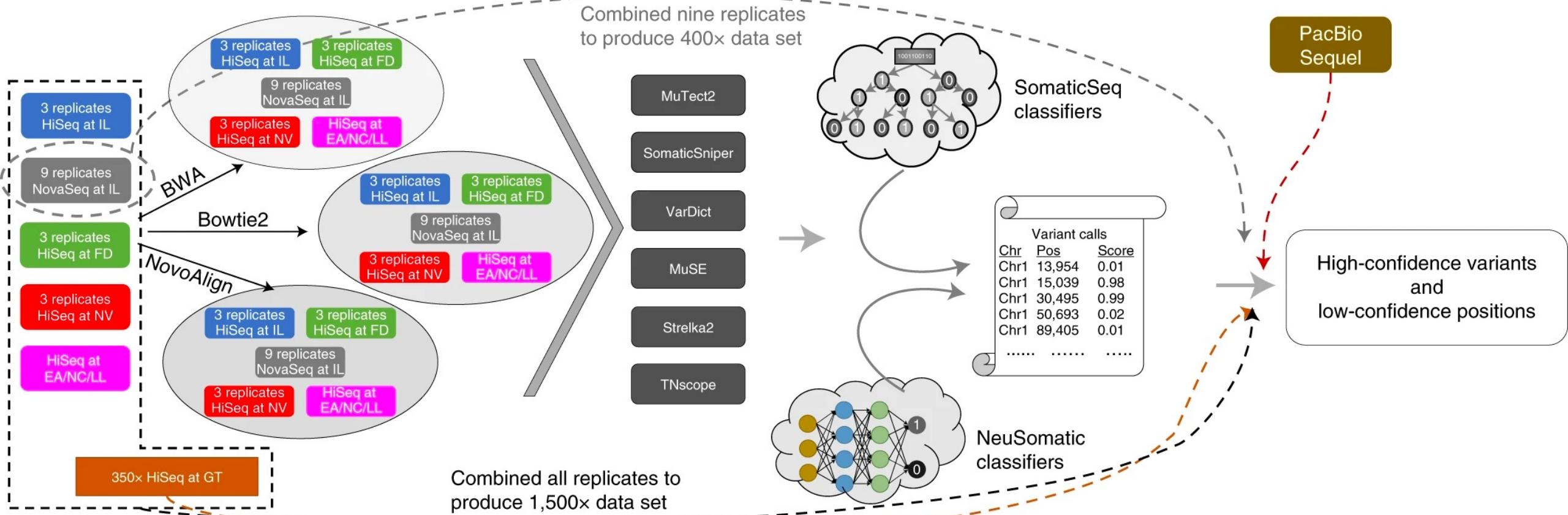
[Nature Biotechnology](#) 39, 1141–1150 (2021) | [Cite this article](#)

7769 Accesses | 14 Citations | 137 Altmetric | [Metrics](#)

# SEQC2 Consortium

| Technology |                        | Platforms            | Sequencing             | Number of reads (coverage) |                     |
|------------|------------------------|----------------------|------------------------|----------------------------|---------------------|
|            |                        |                      |                        | HCC1395                    | HCC1395BL           |
| Discovery  | WGS                    | HiSeq                | 6 centers 12 libraries | 21 billion (750×)          | 21 billion (750×)   |
|            |                        | NovaSeq              | 1 center 9 libraries   | 13 billion (400×)          | 13 billion (400×)   |
| Validation | WGS-tumor content      | HiSeq                | 1 center 3 libraries   | 9.2 billion (350×)         | 9 billion (350×)    |
|            | Tumor-normal mix       | HiSeq                | 1 center 15 libraries  | 46 billion (1,500×)        |                     |
|            | WGS                    | PacBio               | 1 center 1 library     | 20 million (40×)           | 20 million (40×)    |
|            | WES                    | HiSeq                | 6 centers 12 libraries | 3 billion (2,500×)         | 3 billion (2,500×)  |
|            |                        | Ion Torrent          | 1 center 1 library     | 67 million (34×)           | 82 million (47×)    |
|            | AmpliSeq               | MiSeq                | 1 center 1 library     | 25 million (2,900×)        | 22 million (3,300×) |
|            | Microarray             | AffyChip CytoScan HD | 1 center 1 library     | 2.1 million probes         |                     |
|            | Single-cell sequencing | 10x Genomics         | 1 center 1 library     | 1,465 cells                | 983 cells           |

# SEQC2 Consortium



# SEQC2 Consortium - Recomen...

| Process and pipeline               | Recommendations   |
|------------------------------------|---|
| DNA input and library construction | <p>Fragment size</p> <ul style="list-style-type: none"><li>• WGS, 300–600 bp</li><li>• WES, 250–350 bp</li></ul> <p>Fragment method</p> <ul style="list-style-type: none"><li>• Size &gt;250 bp, sonication</li><li>• Size &lt;200 bp, enzyme</li></ul> <p>DNA input</p> <ul style="list-style-type: none"><li>• TruSeq PCR-free, 200 to ~1,000 ng</li><li>• TruSeq-Nano, 10 to ~200 ng</li><li>• Nextera Flex, 1 to ~100 ng</li></ul>            |
| NGS platform choice                | <p>WGS</p> <ul style="list-style-type: none"><li>• More reproducible</li><li>• Limit of detection (LOD) &gt;5% of VAF</li><li>• Read coverage &lt;100× (cost constraints)</li><li>• Best choice if tumor content is high (&gt;50%)</li></ul> <p>WES</p> <ul style="list-style-type: none"><li>• Less reproducible</li><li>• More cost effective</li><li>• High read coverage (&gt;100×)</li><li>• LOD could be as low as 1 to 2% of VAF</li></ul> |

# SEQC2 Consortium

| Process and pipeline                     | Recommendations   |
|--|---|
| Read coverage and quality assurance      | <p>Read coverage</p> <ul style="list-style-type: none"><li>• Tumor content &gt;50%</li><li>◦ 50× (WGS)</li><li>◦ 100× (WES)</li><li>• Tumor content 20 to 50%</li><li>◦ &gt;100× (WGS)</li><li>◦ &gt;200× (WES)</li></ul> <p>QC metrics</p> <ul style="list-style-type: none"><li>• Read redundancy &lt;30%</li><li>• Mappable reads &gt;95%</li><li>• GC content<ul style="list-style-type: none"><li>◦ 40 to 43% (WGS)</li><li>◦ 45 to 48% (WES)</li></ul></li><li>• On-target 65 to ~85%</li><li>• GIV score &lt;1.5</li></ul> |
| Read alignment                           | <p>Alignment quality score range</p> <ul style="list-style-type: none"><li>• Bowtie2, 40 to 50</li><li>• BWA-MEM, 50 to 60</li><li>• NovoAlign, 60 to 70</li></ul> <p>No significant differences between BWA-MEM, Bowtie2 and NovoAlign</p>   |
| Postalignment                            | <p>GATK postalignment processing may not be needed for some mutation callers, such as Strelka2</p> <ul style="list-style-type: none"><li>• Strelka2</li></ul>   |
| Mutation calling                         | <ul style="list-style-type: none"><li>◦ Reliable for WGS</li><li>◦ Sensitive to artifacts with WES</li><li>• MuTect2</li><li>◦ Reliable for both WES and WGS</li><li>◦ Not suitable for WGS with high coverage (200×)</li></ul>   |
| Pipeline construction and change control | <ul style="list-style-type: none"><li>• Detection of cancer mutations is an inherently integrated process</li><li>• Every component and every combination of components is equally important</li><li>• No plug and play</li><li>• Final validation studies and revalidation should be performed using the entire sample-to-result pipeline</li></ul>  |

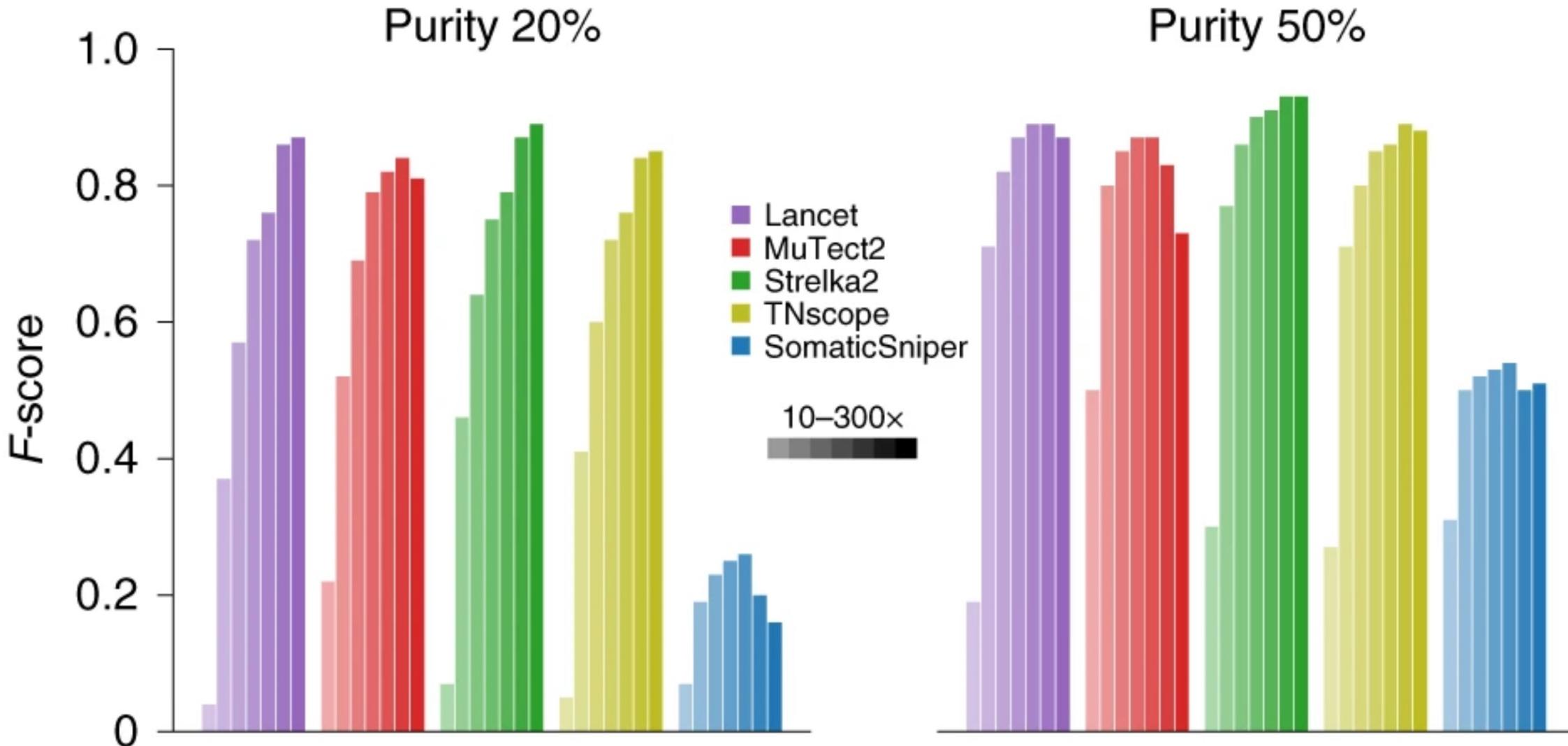
Xiao, W., Ren, L., Chen, Z. *et al.* Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nat Biotechnol* **39**, 1141–1150 (2021). <https://doi.org/10.1038/s41587-021-00994-5>

# SEQC2 Consortium

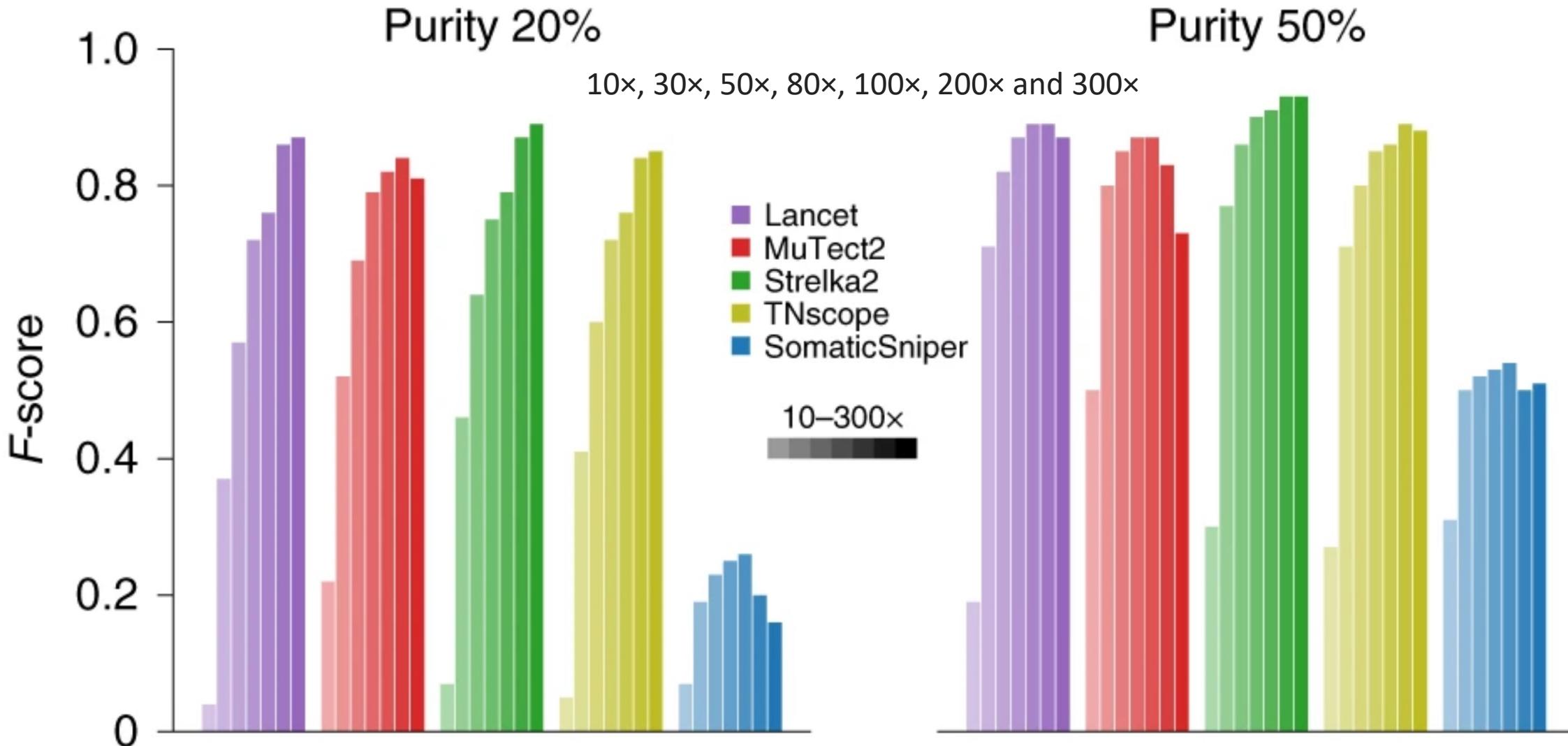
| Process and pipeline                     | Recommendations  |
|--|--|
| Read coverage and quality assurance      | <ul style="list-style-type: none"> <li>Read coverage           <ul style="list-style-type: none"> <li>• Tumor content &gt;50%</li> <li>◦ 50× (WGS) ←</li> <li>◦ 100× (WES) ←</li> <li>• Tumor content 20 to 50%</li> <li>◦ &gt;100× (WGS) ←</li> <li>◦ &gt;200× (WES) ←</li> </ul> </li> <li>QC metrics           <ul style="list-style-type: none"> <li>• Read redundancy &lt;30%</li> <li>• Mappable reads &gt;95%</li> <li>• GC content               <ul style="list-style-type: none"> <li>◦ 40 to 43% (WGS)</li> <li>◦ 45 to 48% (WES)</li> </ul> </li> <li>• On-target 65 to ~85%</li> <li>• GIV score &lt;1.5</li> </ul> </li> </ul> |
| Read alignment                           | <ul style="list-style-type: none"> <li>Alignment quality score range           <ul style="list-style-type: none"> <li>• Bowtie2, 40 to 50</li> <li>• BWA-MEM, 50 to 60</li> <li>• NovoAlign, 60 to 70</li> </ul> </li> <li>No significant differences between BWA-MEM, Bowtie2 and NovoAlign</li> </ul>  |
| Postalignment                            | <ul style="list-style-type: none"> <li>GATK postalignment processing may not be needed for some mutation callers, such as Strelka2</li> </ul>  |
| Mutation calling                         | <ul style="list-style-type: none"> <li>Strelka2</li> <li>◦ Reliable for WGS</li> <li>◦ Sensitive to artifacts with WES</li> <li>MuTect2</li> <li>◦ Reliable for both WES and WGS</li> <li>◦ Not suitable for WGS with high coverage (200×) ←</li> </ul>  |
| Pipeline construction and change control | <ul style="list-style-type: none"> <li>Detection of cancer mutations is an inherently integrated process</li> <li>Every component and every combination of components is equally important</li> <li>No plug and play</li> <li>Final validation studies and revalidation should be performed using the entire sample-to-result pipeline</li> </ul>  |

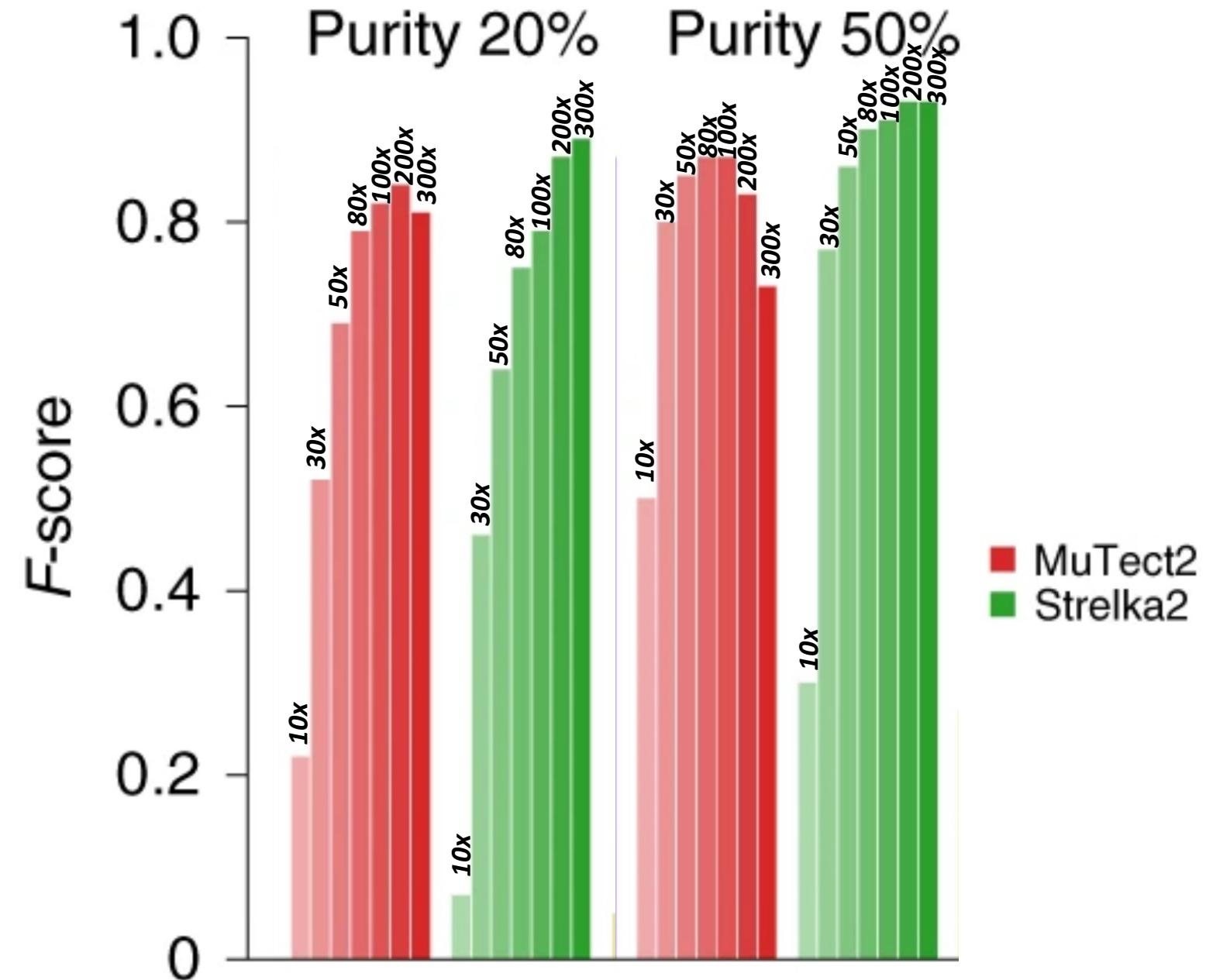
Xiao, W., Ren, L., Chen, Z. *et al.* Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nat Biotechnol* **39**, 1141–1150 (2021). <https://doi.org/10.1038/s41587-021-00994-5>

# SEQC2 Consortium



# SEQC2 Consortium





## SEQC2 - Truthset

- SEQC2 truthset:

<https://www.nature.com/articles/s41587-021-00993-6>

- High Confidence truthset:
  - 39560 SNPs
  - 1922 Indels
- Confidence (bed) file for regions to include.
- 6q ,16p and chrX regions are not part of truth set and confidence region.

# SEQC2 - Truthset

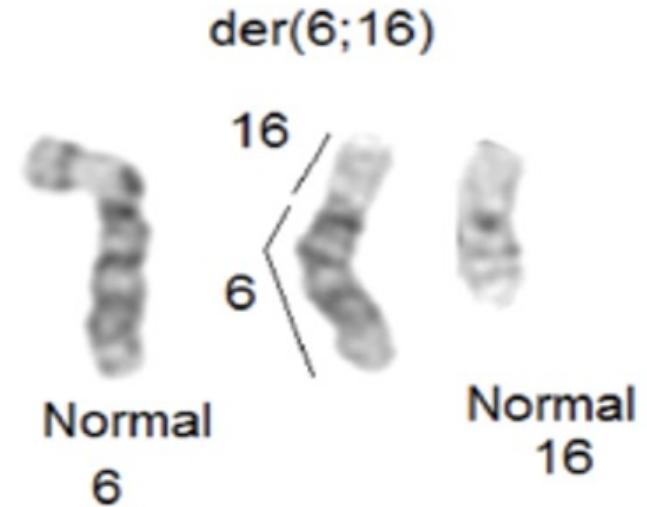
- SEQC2 truthset:

<https://www.nature.com/articles/s41587-021-00993-6>

- High Confidence truthset:

- 39560 SNPs
- 1922 Indels

- Confidence (bed) file for regions to include.
- 6q ,16p and chrX regions are not part of truth set and confidence region.



## SEQC2 - Truthset

- SEQC2 truthset:

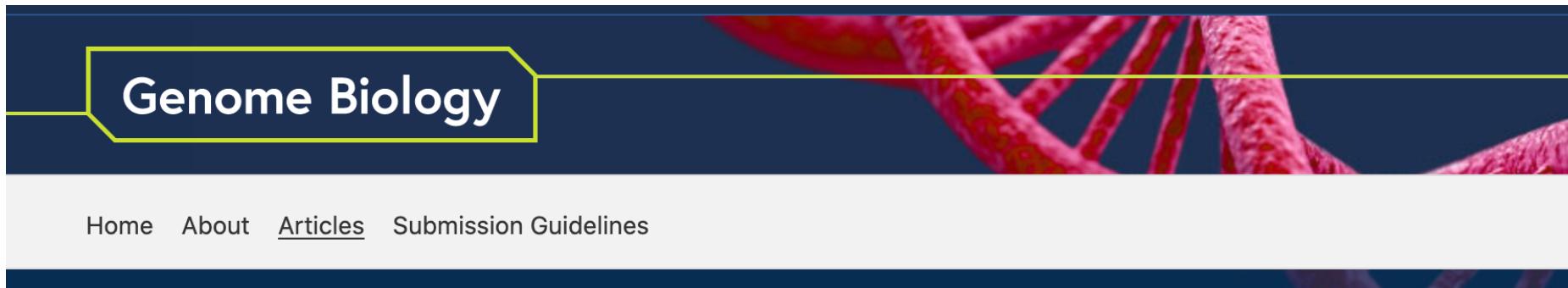
<https://www.nature.com/articles/s41587-021-00993-6>

- High Confidence truthset:

- 39560 SNPs
  - 1922 Indels

- Confidence (bed) file for regions to include.
- 6q ,16p and chrX regions are not part of truth set and confidence region.
- **NC PE 2x150bp sample was used for comparisons ~55x for both T & N.**

# SEQC2 – Tools comparison



Research | [Open Access](#) | [Published: 07 January 2022](#)

## Achieving robust somatic mutation detection with deep learning models derived from reference data sets of a cancer sample

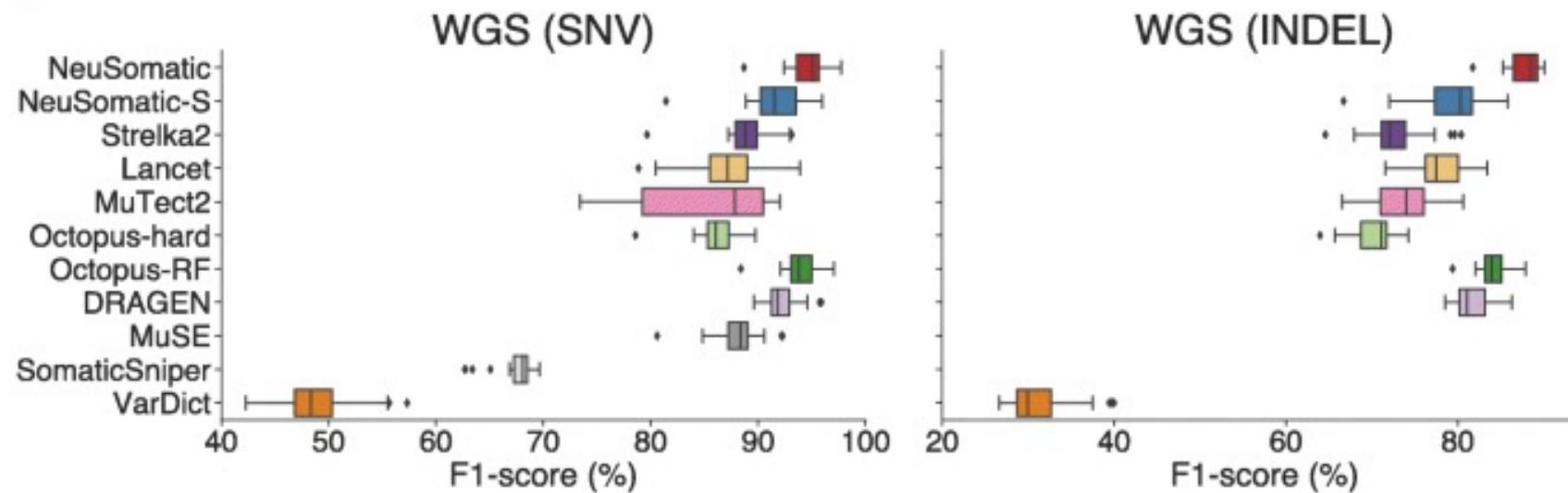
[Sayed Mohammad Ebrahim Sahraeian](#), [Li Tai Fang](#), [Konstantinos Karagiannis](#), [Malcolm Moos](#), [Sean Smith](#), [Luis Santana-Quintero](#), [Chunlin Xiao](#), [Michael Colgan](#), [Huixiao Hong](#), [Marghoob Mohiyuddin](#) & [Wenming Xiao](#)

[Genome Biology](#) 23, Article number: 12 (2022) | [Cite this article](#)

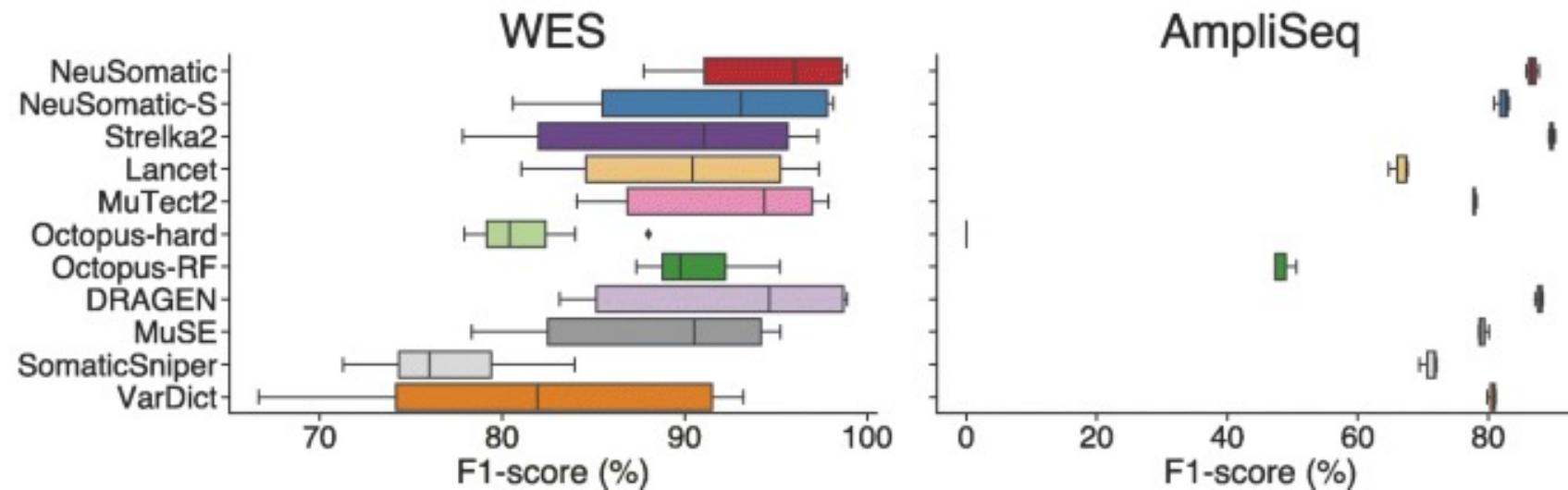
2481 Accesses | 11 Altmetric | [Metrics](#)

# SEQC2 – Tools comparison

a

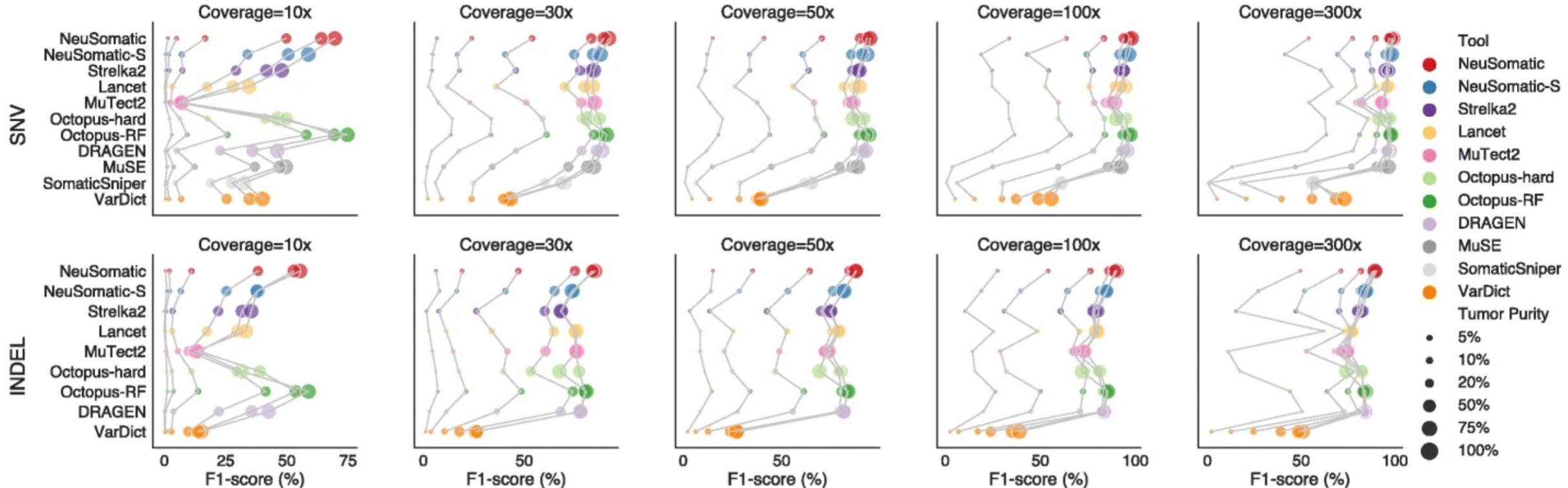


b

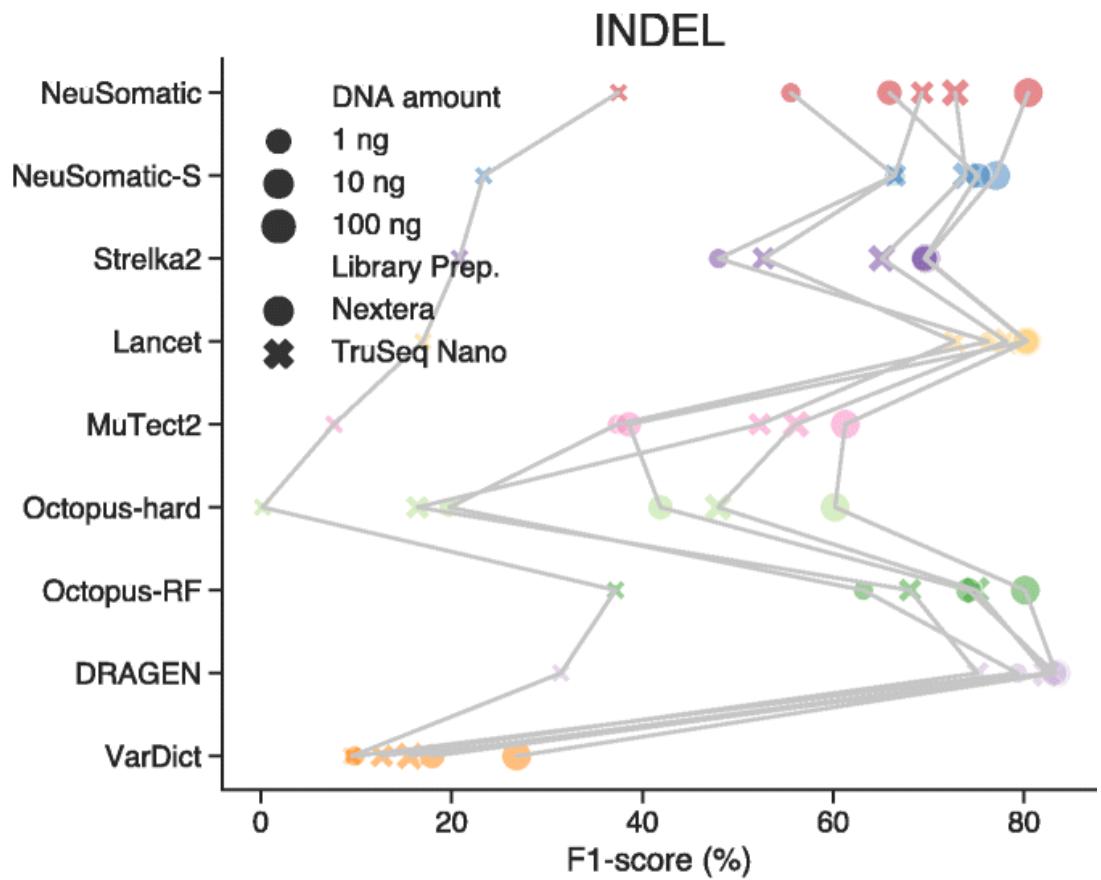
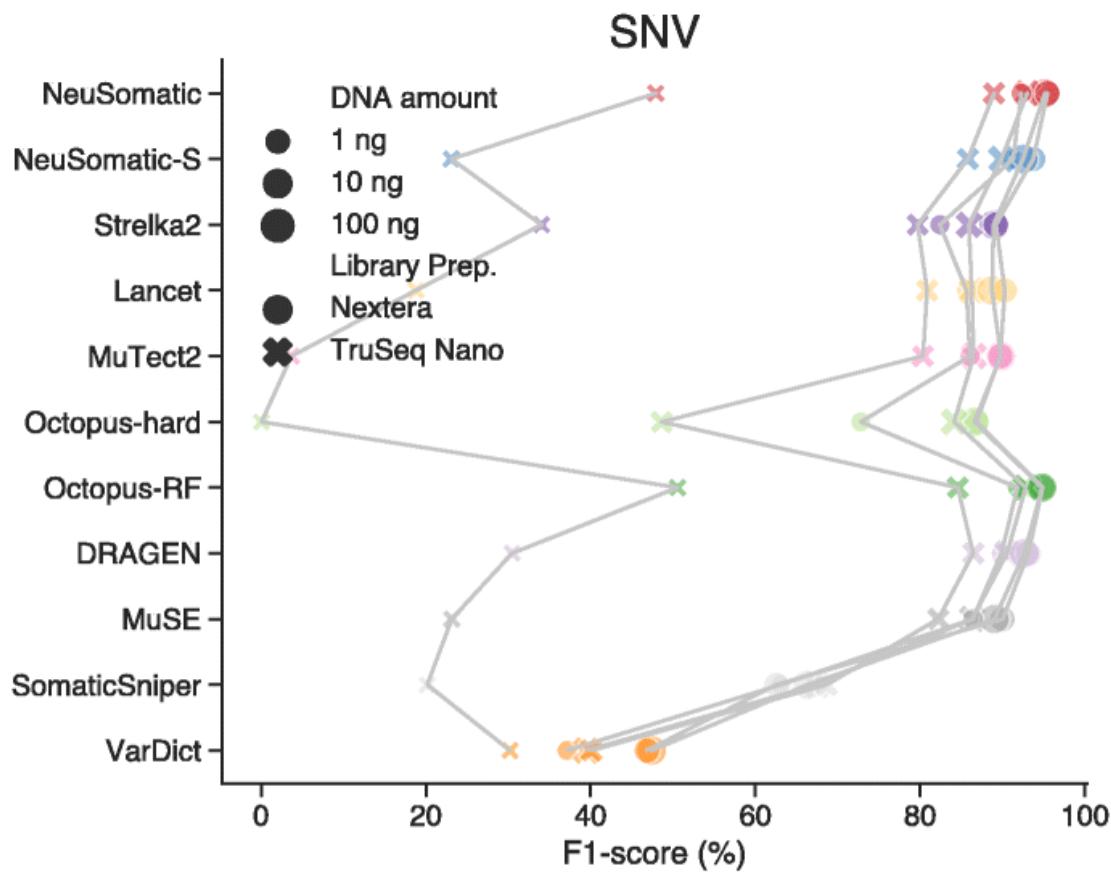


# SEQC2 – Tools comparison

a

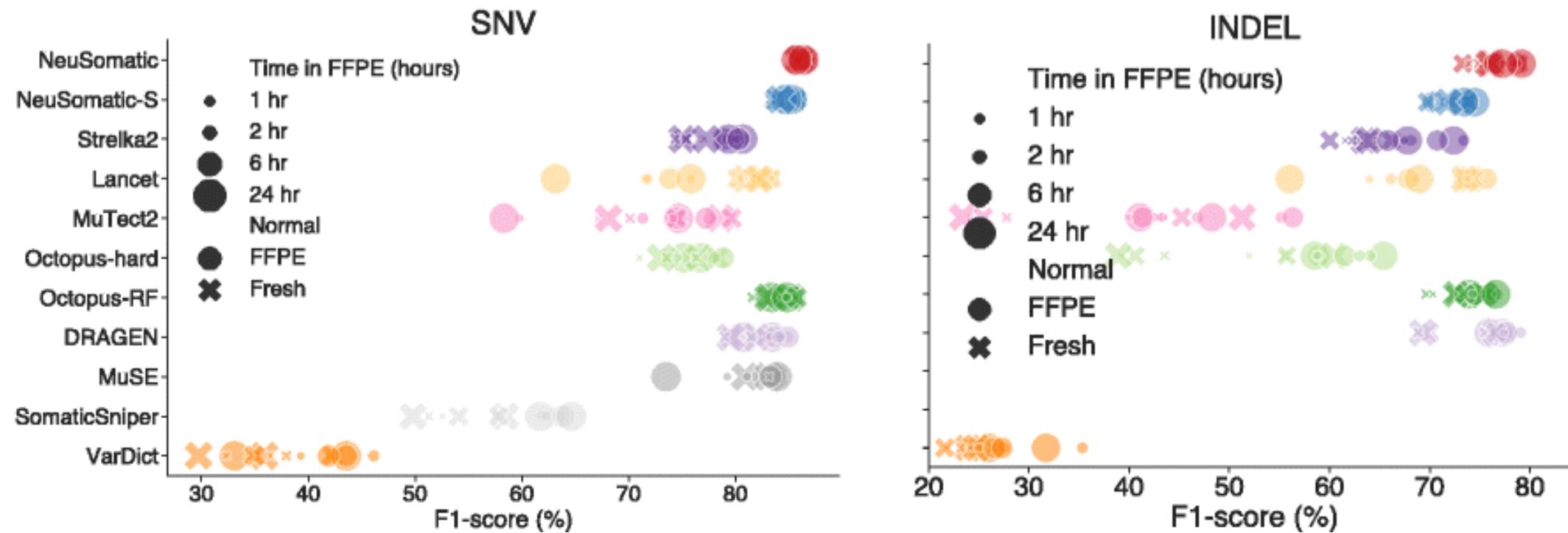


# SEQC2 – Tools comparison



# SEQC2 – Tools comparison

C



## SEQC2 – Tools comparison

- Only shown F1 score which does not give correct results for all the aspects i.e. for Somatic variant stools we know Precision would be higher than Recall!
- After creating High Confidence truthset and confidence region file, both high & mid confidante truthset. Confidence region file was not used.
- Purpose for inhouse comparison to know which tool works the best in best possible conditions .

# Tools & Comparison

- Dragen – v3.7,3.8,3.9 & 3.10
  - Mutect2 – v4.2.0
  - Stelka: v2.9.10
  - Octopus: v7.2 & v7.4
  - Muse: r1
  - Somatic Sniper: latest
  - Sage (hmftools) : v2.8
  - Lancet: v1.1.0
  - LoFreq – v2.1.5
- 
- Variant calling comparison was done by using bam file from Dragen v.3.8
  - GDC somatic PON was used for comparison.

# Tools Used

- Dragen, Strelka, Somatic Sniper & Mutect2 are widely used somatic tools.
- Sage : hmftools
  - Same group who made Purple
  - Group is known for making tools for clinical research.
  - Sage is extremely fast - ~30 mins to do variant calling using 36 cores.
  - Uses their own PON with custom tags to tag and filter variants
- Muse:
  - Published in 2016 but recently (2022) updated.
  - Tested after finding that Baylor clinical sequencing uses it.
  - Also part of GDC somatic workflow.
  - Use dbsnp to filter germline calls.
- Strelka:
  - Use “--callRegions”
    - With Confidence regions : 2030 sec
    - With chr1to22, X & Y as call regions : 2090 sec
    - Without CallRegions : 74086 sec

# Tools Used

- Octopus:
  - Lasted variant caller that can do all
    - Germline, Trio, Somatic, Somatic T-only, population, Polyclone, Singcell etc....
    - 2 models to run somatic/germline – default & Random Forest
    - Error models for PCR/PCR-Free & for Hiseq2000/2500/4000/X10, Novaseq/Novaseq PCR & BIGSEQ5000
- Lancet
  - From NY Genome center - uses a localized micro-assembly strategy to detect somatic mutation
  - Can be run per chr only
  - No indels

# Somatic Pair - SNPs

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.10      | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |
| Lancet            | SNPS        | PASS          | 34804            | 39560            | 34451     | 353       | 5109      | 0.9899           | 0.8709        | 0.92655   |
| Muse              | SNPS        | PASS          | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |
| Mutect2_v4.2.0    | SNPS        | PASS          | 33980            | 39560            | 33656     | 324       | 5904      | 0.9905           | 0.8508        | 0.91531   |
| Sage_v3.8         | SNPS        | PASS          | 32199            | 39560            | 32132     | 67        | 7428      | 0.9979           | 0.8122        | 0.89555   |
| Somatic Sniper    | SNPS        | ALL           | 46721            | 39560            | 39560     | 15003     | 7842      | 0.8467           | 0.8346        | 0.77595   |
| Strelka           | SNPS        | PASS          | 37706            | 39560            | 36511     | 1195      | 3029      | 0.9683           | 0.9234        | 0.94532   |
| Octopus_v7.4 + RF | SNPS        | PASS          | 37456            | 39560            | 36940     | 516       | 2620      | 0.9862           | 0.9338        | 0.95928   |
| LoFreq2.1.5       | SNPS        | PASS          | 27411            | 39560            | 27364     | 47        | 12196     | 0.9983           | 0.6917        | 0.81719   |

# Somatic Pair - SNPs

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.10      | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |
| Lancet            | SNPS        | PASS          | 34804            | 39560            | 34451     | 353       | 5109      | 0.9899           | 0.8709        | 0.92655   |
| Muse              | SNPS        | PASS          | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |
| Mutect2_v4.2.0    | SNPS        | PASS          | 33980            | 39560            | 33656     | 324       | 5904      | 0.9905           | 0.8508        | 0.91531   |
| Sage_v3.8         | SNPS        | PASS          | 32199            | 39560            | 32132     | 67        | 7428      | 0.9979           | 0.8122        | 0.89555   |
| Somatic Sniper    | SNPS        | ALL           | 46721            | 39560            | 39560     | 15003     | 7842      | 0.8467           | 0.8346        | 0.77595   |
| Strelka           | SNPS        | PASS          | 37706            | 39560            | 36511     | 1195      | 3029      | 0.9683           | 0.9234        | 0.94532   |
| Octopus_v7.4 + RF | SNPS        | PASS          | 37456            | 39560            | 36940     | 516       | 2620      | 0.9862           | 0.9338        | 0.95928   |
| LoFreq2.1.5       | SNPS        | PASS          | 27411            | 39560            | 27364     | 47        | 12196     | 0.9983           | 0.6917        | 0.81719   |

# Somatic Pair - SNPs

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.10      | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |
| Lancet            | SNPS        | PASS          | 34804            | 39560            | 34451     | 353       | 5109      | 0.9899           | 0.8709        | 0.92655   |
| Muse              | SNPS        | PASS          | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |
| Mutect2_v4.2.0    | SNPS        | PASS          | 33980            | 39560            | 33656     | 324       | 5904      | 0.9905           | 0.8508        | 0.91531   |
| Sage_v3.8         | SNPS        | PASS          | 32199            | 39560            | 32132     | 67        | 7428      | 0.9979           | 0.8122        | 0.89555   |
| Somatic Sniper    | SNPS        | ALL           | 46721            | 39560            | 39560     | 15003     | 7842      | 0.8467           | 0.8346        | 0.77595   |
| Strelka           | SNPS        | PASS          | 37706            | 39560            | 36511     | 1195      | 3029      | 0.9683           | 0.9234        | 0.94532   |
| Octopus_v7.4 + RF | SNPS        | PASS          | 37456            | 39560            | 36940     | 516       | 2620      | 0.9862           | 0.9338        | 0.95928   |
| LoFreq2.1.5       | SNPS        | PASS          | 27411            | 39560            | 27364     | 47        | 12196     | 0.9983           | 0.6917        | 0.81719   |



# Somatic Pair - SNPs

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.10      | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |
| Lancet            | SNPS        | PASS          | 34804            | 39560            | 34451     | 353       | 5109      | 0.9899           | 0.8709        | 0.92655   |
| Muse              | SNPS        | PASS          | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |
| Mutect2_v4.2.0    | SNPS        | PASS          | 33980            | 39560            | 33656     | 324       | 5904      | 0.9905           | 0.8508        | 0.91531   |
| Sage_v3.8         | SNPS        | PASS          | 32199            | 39560            | 32132     | 67        | 7428      | 0.9979           | 0.8122        | 0.89555   |
| Somatic Sniper    | SNPS        | ALL           | 46721            | 39560            | 39560     | 15003     | 7842      | 0.8467           | 0.8346        | 0.77595   |
| Strelka           | SNPS        | PASS          | 37706            | 39560            | 36511     | 1195      | 3029      | 0.9683           | 0.9234        | 0.94532   |
| Octopus_v7.4 + RF | SNPS        | PASS          | 37456            | 39560            | 36940     | 516       | 2620      | 0.9862           | 0.9338        | 0.95928   |
| LoFreq2.1.5       | SNPS        | PASS          | 27411            | 39560            | 27364     | 47        | 12196     | 0.9983           | 0.6917        | 0.81719   |

| <i>Tool</i>   | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|---------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.8.4 | SNPS        | PASS          | 36137            | 39560            | 35888     | 249       | 3672      | 0.9931           | 0.9072        | 0.94820   |
| Dragen_v3.9   | SNPS        | PASS          | 36485            | 39560            | 36208     | 277       | 3352      | 0.9924           | 0.9153        | 0.95228   |
| Dragen_v3.7   | SNPS        | PASS          | 36120            | 39560            | 35859     | 261       | 3701      | 0.9928           | 0.9064        | 0.94765   |
| Dragen_v3.10  | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |

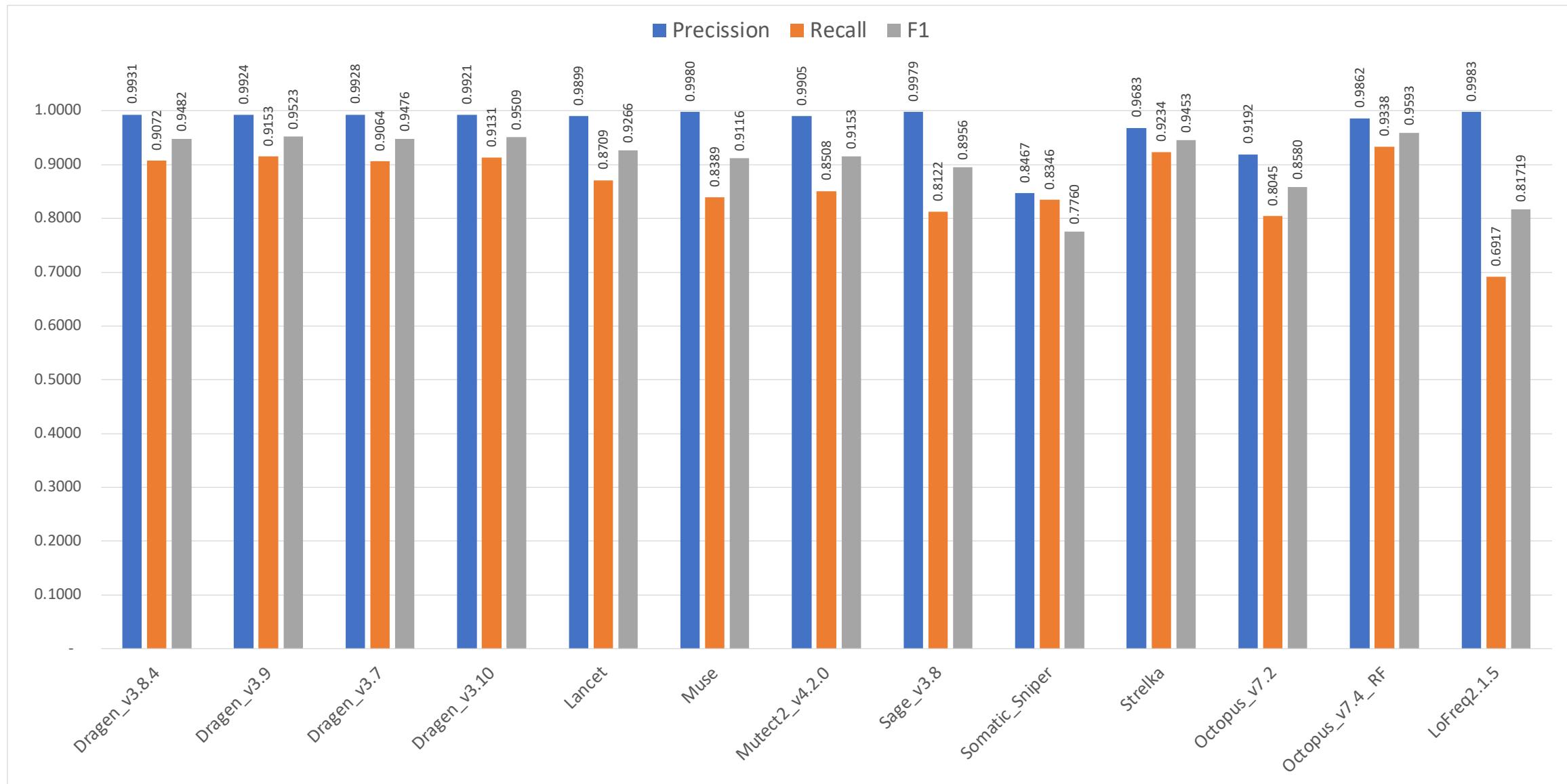
# Somatic Pair - SNPs

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.10      | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |
| Lancet            | SNPS        | PASS          | 34804            | 39560            | 34451     | 353       | 5109      | 0.9899           | 0.8709        | 0.92655   |
| Muse              | SNPS        | PASS          | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |
| Mutect2_v4.2.0    | SNPS        | PASS          | 33980            | 39560            | 33656     | 324       | 5904      | 0.9905           | 0.8508        | 0.91531   |
| Sage_v3.8         | SNPS        | PASS          | 32199            | 39560            | 32132     | 67        | 7428      | 0.9979           | 0.8122        | 0.89555   |
| Somatic Sniper    | SNPS        | ALL           | 46721            | 39560            | 39560     | 15003     | 7842      | 0.8467           | 0.8346        | 0.77595   |
| Strelka           | SNPS        | PASS          | 37706            | 39560            | 36511     | 1195      | 3029      | 0.9683           | 0.9234        | 0.94532   |
| Octopus_v7.4 + RF | SNPS        | PASS          | 37456            | 39560            | 36940     | 516       | 2620      | 0.9862           | 0.9338        | 0.95928   |
| LoFreq2.1.5       | SNPS        | PASS          | 27411            | 39560            | 27364     | 47        | 12196     | 0.9983           | 0.6917        | 0.81719   |

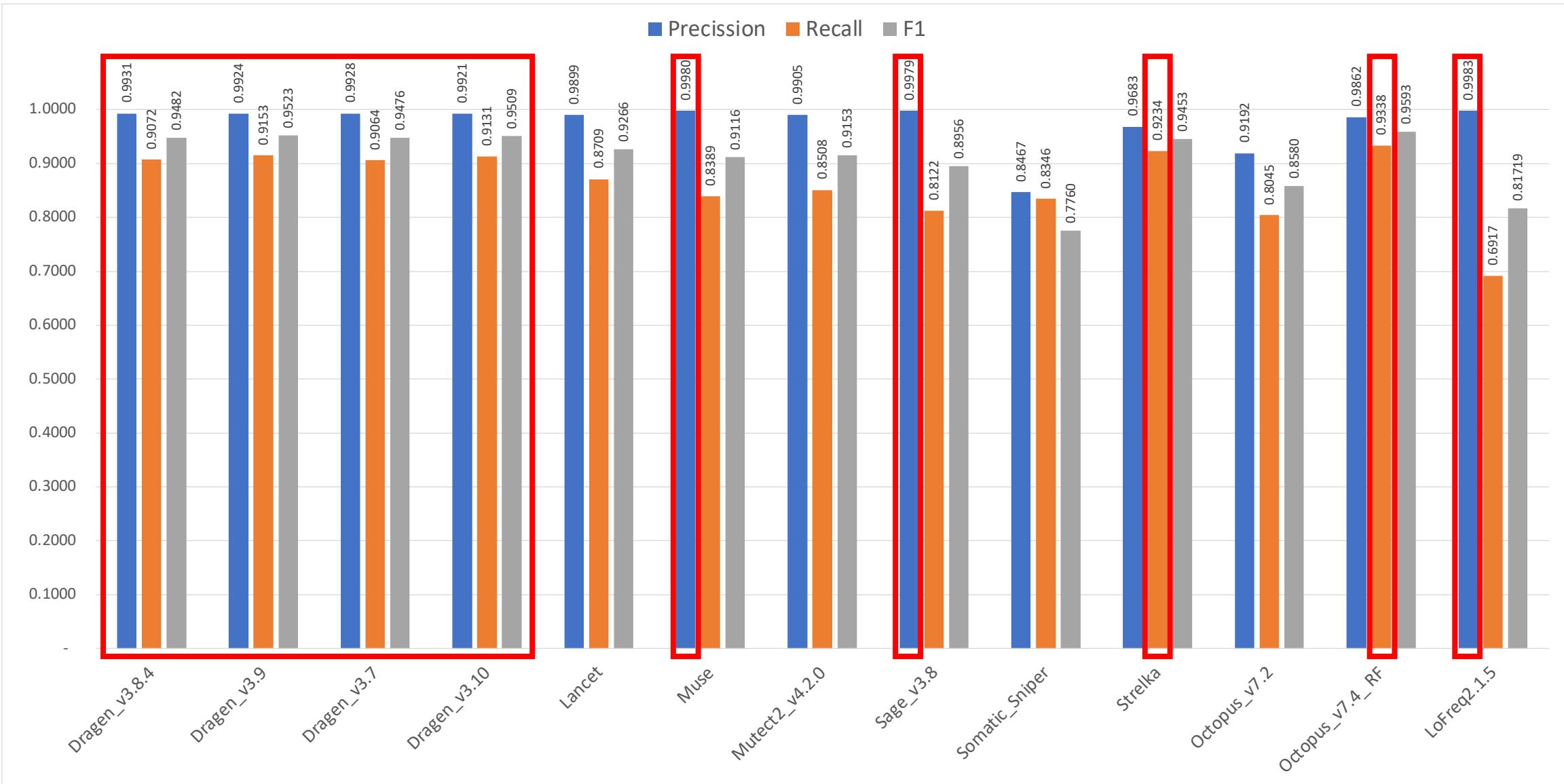
| <i>Tool</i>   | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|---------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.8.4 | SNPS        | PASS          | 36137            | 39560            | 35888     | 249       | 3672      | 0.9931           | 0.9072        | 0.94820   |
| Dragen_v3.9   | SNPS        | PASS          | 36485            | 39560            | 36208     | 277       | 3352      | 0.9924           | 0.9153        | 0.95228   |
| Dragen_v3.7   | SNPS        | PASS          | 36120            | 39560            | 35859     | 261       | 3701      | 0.9928           | 0.9064        | 0.94765   |
| Dragen_v3.10  | SNPS        | PASS          | 36409            | 39560            | 36121     | 288       | 3439      | 0.9921           | 0.9131        | 0.95094   |

| <i>Tool</i> | <i>Type</i> | <i>R_Type</i>  | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------|-------------|----------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Muse        | SNPS        | PASS + T1      | 36408            | 39560            | 35921     | 487       | 3639      | 0.9866           | 0.9080        | 0.94569   |
| Muse        | SNPS        | PASS + T1 + T5 | 37591            | 39560            | 36408     | 1183      | 3152      | 0.9685           | 0.9203        | 0.94381   |
| Muse        | SNPS        | PASS           | 33253            | 39560            | 33188     | 65        | 6372      | 0.9980           | 0.8389        | 0.91160   |

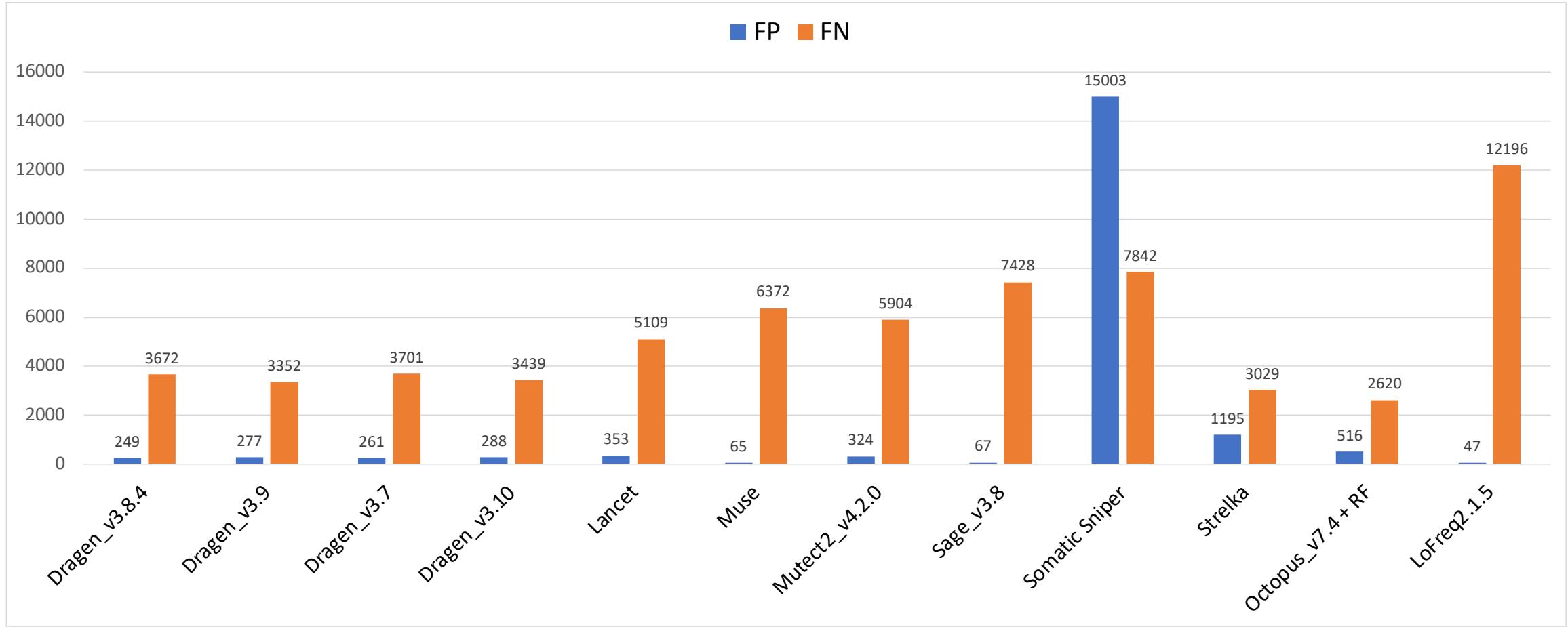
# Somatic Pair - SNPs



# Somatic Pair - SNPs

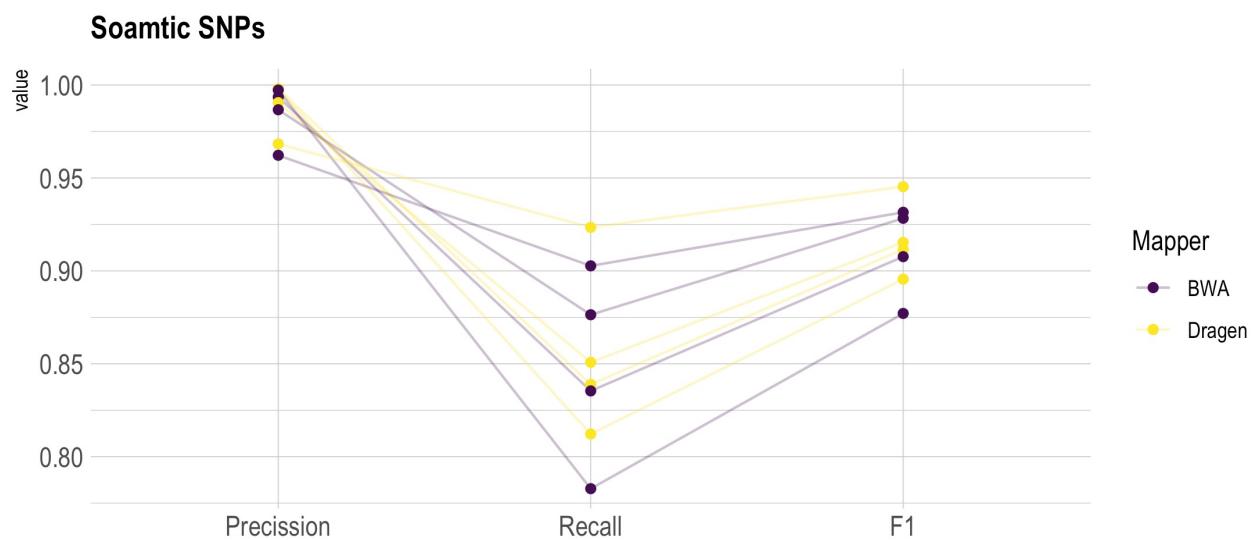


# Somatic Pair - SNPs



# Dragen vs BWA ALT

| Tool           | Mapper | Type | R_Type | Total - Q | Total - T | TP    | FP   | FN   | Precision | Recall | F1      |
|----------------|--------|------|--------|-----------|-----------|-------|------|------|-----------|--------|---------|
| Muse           | Dragen | SNPS | PASS   | 33253     | 39560     | 33188 | 65   | 6372 | 0.9980    | 0.8389 | 0.91160 |
| Muse           | BWA    | SNPS | PASS   | 33262     | 39560     | 33048 | 214  | 6512 | 0.9936    | 0.8354 | 0.90764 |
| Mutect2_v4.2.0 | Dragen | SNPS | PASS   | 33980     | 39560     | 33656 | 324  | 5904 | 0.9905    | 0.8508 | 0.91531 |
| Mutect2_v4.2.0 | BWA    | SNPS | PASS   | 35138     | 39560     | 34671 | 467  | 4889 | 0.9867    | 0.8764 | 0.92830 |
| Sage_v3.8      | Dragen | SNPS | PASS   | 32199     | 39560     | 32132 | 67   | 7428 | 0.9979    | 0.8122 | 0.89555 |
| Sage_v3.8      | BWA    | SNPS | PASS   | 31052     | 39560     | 30967 | 85   | 8593 | 0.9973    | 0.7828 | 0.87710 |
| Strelka        | Dragen | SNPS | PASS   | 37706     | 39560     | 36511 | 1195 | 3029 | 0.9683    | 0.9234 | 0.94532 |
| Strelka_Sarek  | BWA    | SNPS | PASS   | 37114     | 39560     | 35712 | 1402 | 3848 | 0.9622    | 0.9027 | 0.93153 |



# Somatic Pair – SNPs – Intersect of callers

| Tool                    | Type | R_Type | Total - Q | Total - T | TP    | FP  | FN   | Precision | Recall | F1     |
|-------------------------|------|--------|-----------|-----------|-------|-----|------|-----------|--------|--------|
| any2_DraStrMut          | SNPS | PASS   | 36071     | 39560     | 35966 | 105 | 3594 | 0.9971    | 0.9092 | 0.9511 |
| any2_DraStrSag          | SNPS | PASS   | 36142     | 39560     | 36030 | 112 | 3530 | 0.9969    | 0.9108 | 0.9519 |
| any2_DraStrSagMut       | SNPS | PASS   | 36492     | 39560     | 36353 | 139 | 3207 | 0.9962    | 0.9189 | 0.9560 |
| any2_StrSagMut          | SNPS | PASS   | 34647     | 39560     | 34555 | 92  | 5005 | 0.9973    | 0.8735 | 0.9313 |
| any2_DraLanMutMusStr    | SNPS | PASS   | 36969     | 39560     | 36791 | 187 | 2769 | 0.9952    | 0.9300 | 0.9614 |
| any2_DraLanMutMusSagStr | SNPS | PASS   | 37020     | 39560     | 36820 | 200 | 2740 | 0.9946    | 0.9307 | 0.9616 |
| Any2_LanMutMusSagStr    | SNPS | PASS   | 36728     | 39560     | 36577 | 151 | 2983 | 0.9959    | 0.9246 | 0.9589 |
| Any2_LanMutMusStr       | SNPS | PASS   | 36658     | 39560     | 36528 | 130 | 3032 | 0.9965    | 0.9234 | 0.9585 |
| any2_DraMutMusStrOct    | SNPS | PASS   | 37304     | 39560     | 37096 | 208 | 2464 | 0.9944    | 0.9377 | 0.9652 |
| any2_MutMusStrOct       | SNPS | PASS   | 37104     | 39560     | 36945 | 159 | 2615 | 0.9957    | 0.9339 | 0.9638 |

Dra – Dragen

Mus – Muse

Mut – Mutect2

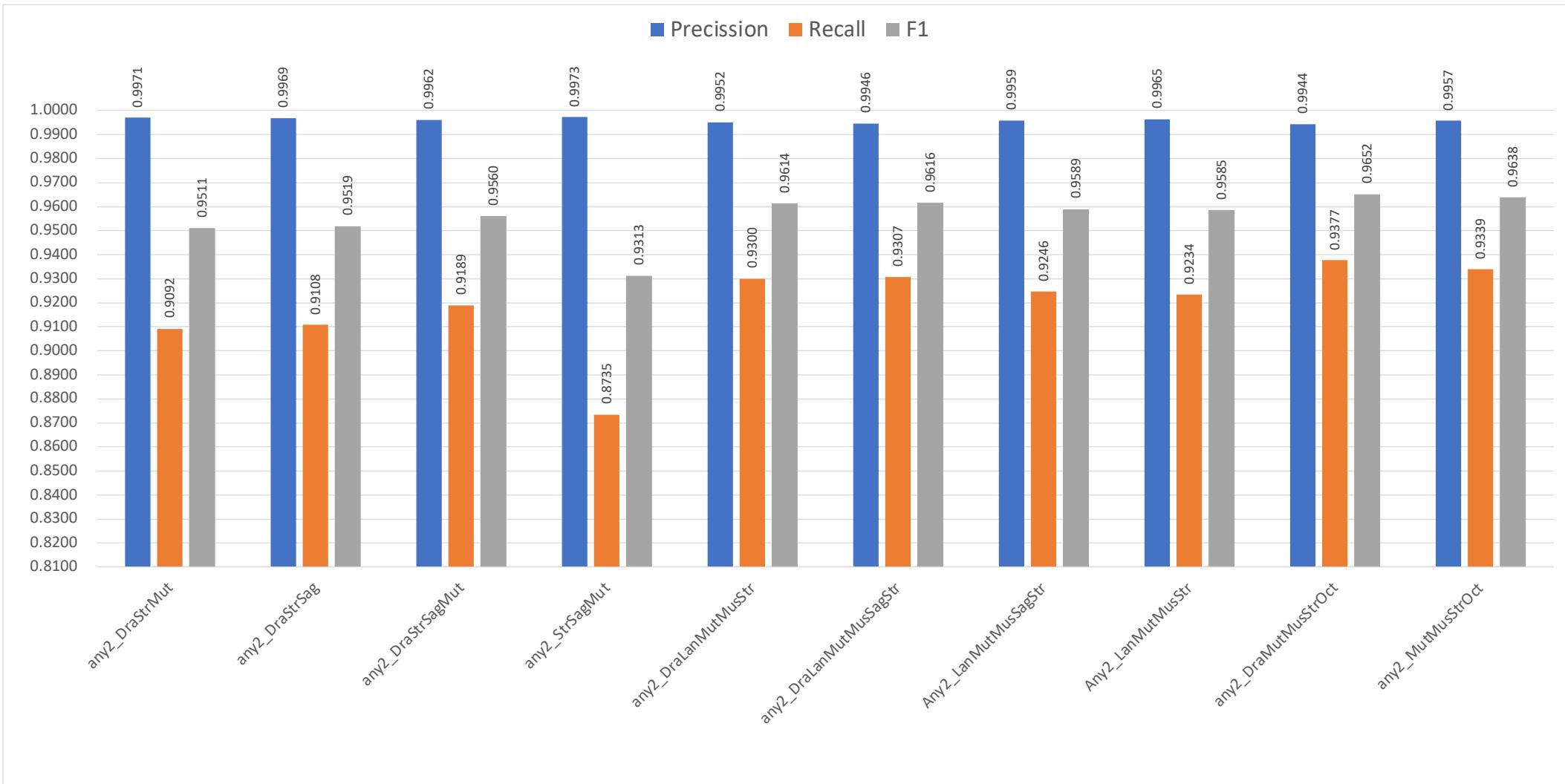
Oct – Octopus RF

Sag – Sage

Str – Strelka

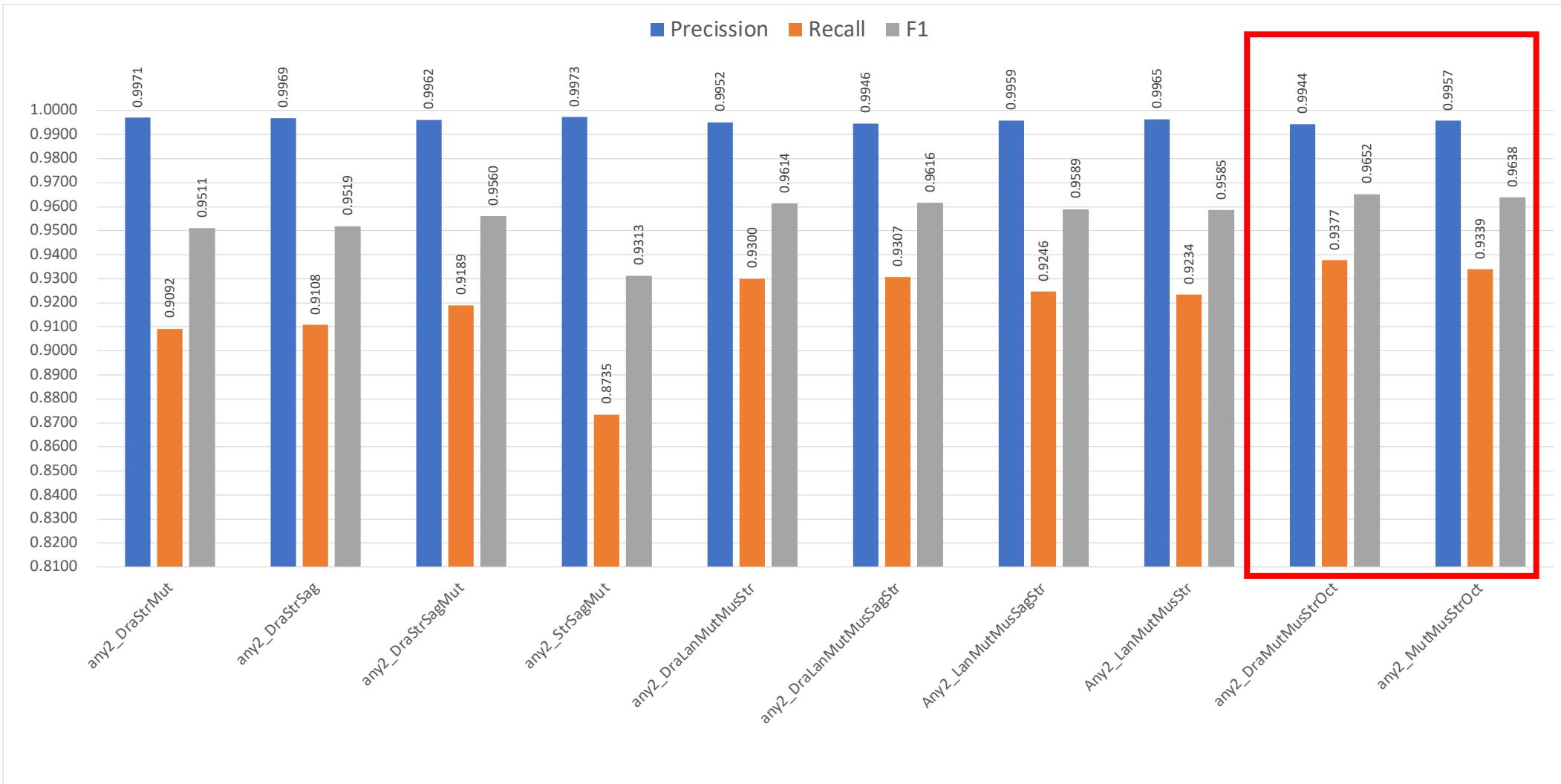
# Somatic Pair – SNPs – Intersect of callers

Dra – Dragen  
Mus – Muse  
Mut – Mutect2  
Oct – Octopus RF  
Sag – Sage  
Str – Strelka



# Somatic Pair – SNPs – Intersect of callers

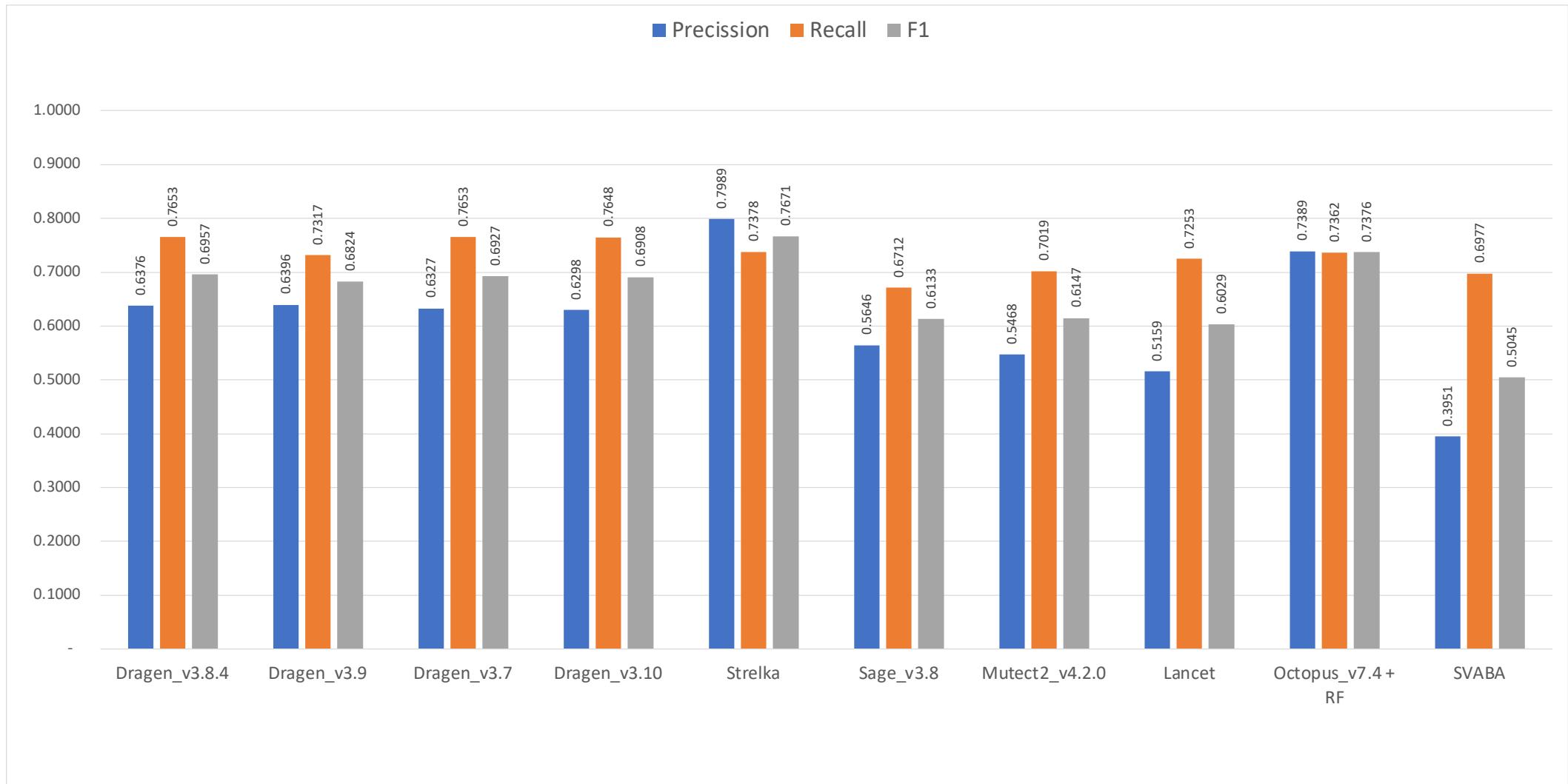
Dra – Dragen  
Mus – Muse  
Mut – Mutect2  
Oct – Octopus RF  
Sag – Sage  
Str – Strelka



# Somatic Pair - Indels

| <i>Tool</i>       | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-------------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.8.4     | INDELS      | PASS          | 2307             | 1922             | 1471      | 836       | 451       | 0.6376           | 0.7653        | 0.6957    |
| Dragen_v3.9       | INDELS      | PASS          | 2303             | 1922             | 1473      | 831       | 540       | 0.6396           | 0.7317        | 0.6824    |
| Dragen_v3.7       | INDELS      | PASS          | 2325             | 1922             | 1471      | 854       | 451       | 0.6327           | 0.7653        | 0.6927    |
| Dragen_v3.10      | INDELS      | PASS          | 2334             | 1922             | 1470      | 864       | 452       | 0.6298           | 0.7648        | 0.6908    |
| Strelka           | INDELS      | PASS          | 1775             | 1922             | 1418      | 357       | 504       | 0.7989           | 0.7378        | 0.7671    |
| Sage_v3.8         | INDELS      | PASS          | 2285             | 1922             | 1290      | 995       | 632       | 0.5646           | 0.6712        | 0.6133    |
| Mutect2_v4.2.0    | INDELS      | PASS          | 2467             | 1922             | 1349      | 1118      | 573       | 0.5468           | 0.7019        | 0.6147    |
| Lancet            | INDELS      | PASS          | 2702             | 1922             | 1394      | 1308      | 528       | 0.5159           | 0.7253        | 0.6029    |
| Octopus_v7.4 + RF | INDELS      | PASS          | 1915             | 1922             | 1415      | 500       | 507       | 0.7389           | 0.7362        | 0.7376    |
| SVABA             | INDELS      | PASS          | 3394             | 1922             | 1341      | 2053      | 581       | 0.3951           | 0.6977        | 0.5045    |

# Somatic Pair - Indels



# Somatic Pair – Indels – Intersect of callers

| Tool              | Type   | R_Type | Total - Q | Total - T | TP   | FP  | FN  | Precision | Recall | F1     |
|-------------------|--------|--------|-----------|-----------|------|-----|-----|-----------|--------|--------|
| any2_DraStrMut    | INDELS | PASS   | 1922      | 1922      | 1453 | 469 | 469 | 0.7560    | 0.7560 | 0.7560 |
| any2_DraSrtSagMut | INDELS | PASS   | 2348      | 1922      | 1482 | 866 | 440 | 0.6312    | 0.7711 | 0.6941 |
| any2_DraStrSagMut | INDELS | PASS   | 1655      | 1922      | 1439 | 216 | 483 | 0.8695    | 0.7487 | 0.8046 |
| any2_StrSagMut    | INDELS | PASS   | 1909      | 1922      | 1399 | 510 | 523 | 0.7328    | 0.7279 | 0.7304 |

Dra – Dragen

Mut – Mutect2

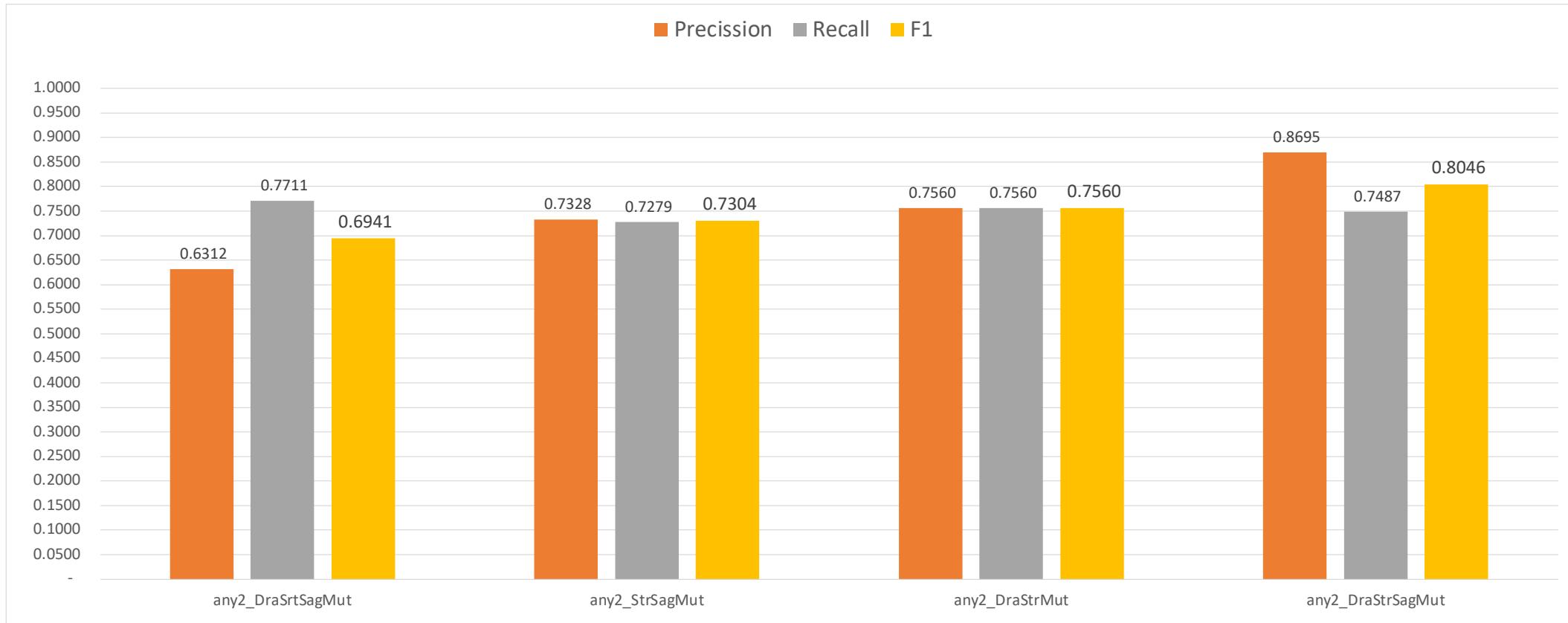
Oct – Octopus RF

Sag – Sage

Str – Strelka

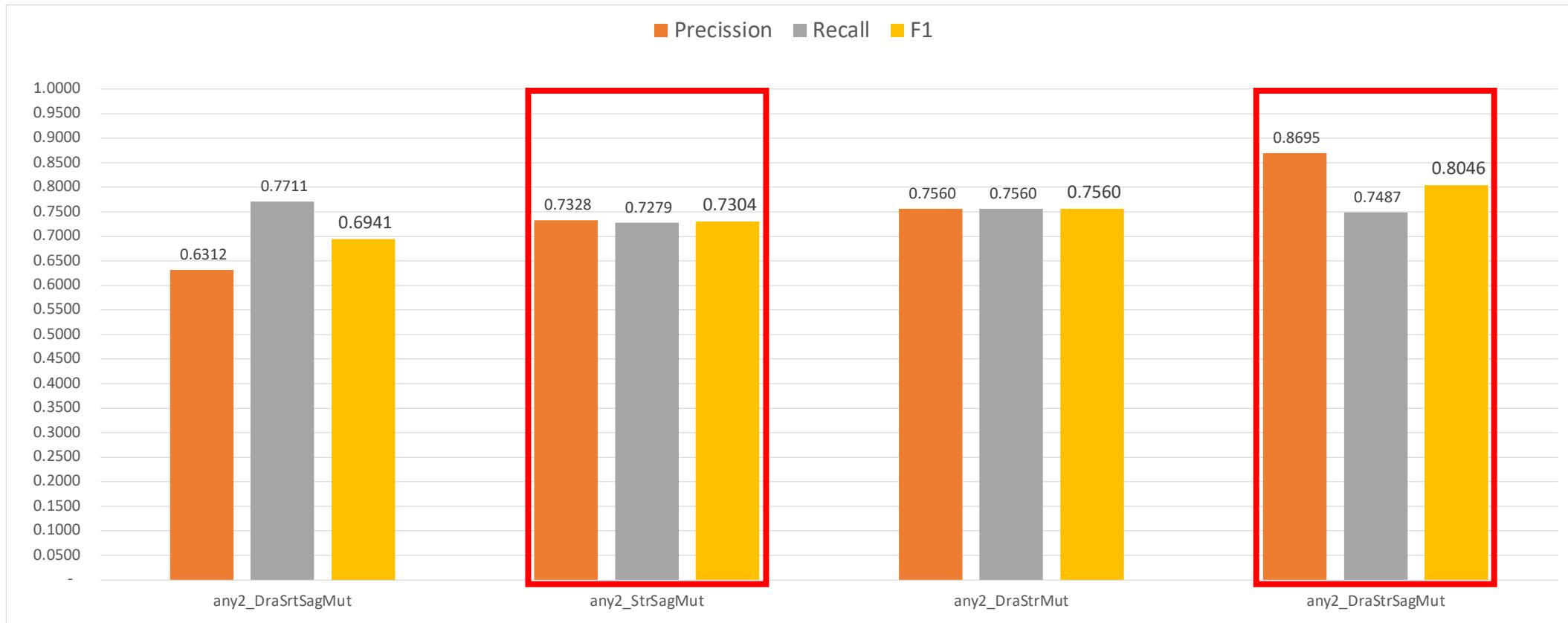
# Somatic Pair – Indels – Intersect of callers

Dra – Dragen  
Mut – Mutect2  
Oct – Octopus RF  
Sag – Sage  
Str – Strelka



# Somatic Pair – Indels – Intersect of callers

Dra – Dragen  
Mut – Mutect2  
Oct – Octopus RF  
Sag – Sage  
Str – Strelka



## Somatic Tumor-only

Tumor only variant calling is **performed on a tumor sample with no paired normal at the request of the research group**. This method takes advantage of the normal cell contamination that is present in most tumor samples.

## Somatic Tumor-only

Tumor only variant calling is **performed on a tumor sample with no paired normal at the request of the research group**. This method takes advantage of the normal cell contamination that is present in most tumor samples.

| <i>Tool</i>     | <i>Type</i> | <i>R_Type</i> | <i>Total - Q</i> | <i>Total - T</i> | <i>TP</i> | <i>FP</i> | <i>FN</i> | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|-----------------|-------------|---------------|------------------|------------------|-----------|-----------|-----------|------------------|---------------|-----------|
| Dragen_v3.7     | SNPS        | PASS          | 241631           | 39560            | 35990     | 205641    | 3570      | 0.1489           | 0.9098        | 0.2560    |
| Dragen_v3.8.4   | SNPS        | PASS          | 238448           | 39560            | 36026     | 202422    | 3534      | 0.1511           | 0.9107        | 0.2592    |
| Dragen_v3.9     | SNPS        | PASS          | 239362           | 39560            | 36387     | 202975    | 2137      | 0.1520           | 0.9445        | 0.2619    |
| Dragen_v3.10    | SNPS        | PASS          | 50108            | 39560            | 35993     | 14115     | 3567      | 0.7183           | 0.9098        | 0.8028    |
| Sage_v3.8       | SNPS        | PASS          | 68504            | 39560            | 31844     | 36660     | 7716      | 0.4648           | 0.8050        | 0.5894    |
| Mutect2_v4.2.0  | SNPS        | PASS          | 57304            | 39560            | 24757     | 32547     | 14803     | 0.4320           | 0.6258        | 0.5112    |
| Any2_DraMutSag  | SNPS        | PASS          | 80974            | 39560            | 34974     | 46000     | 4586      | 0.4319           | 0.8841        | 0.5803    |
| Common_DraMut   | SNPS        | PASS          | 48954            | 39560            | 24342     | 24612     | 15218     | 0.4972           | 0.6153        | 0.5500    |
| Common_MutSag   | SNPS        | PASS          | 33237            | 39560            | 21068     | 12169     | 18492     | 0.6339           | 0.5326        | 0.5788    |
| Octopus_v7.4_RF | SNPS        | PASS          | 44374            | 39560            | 12694     | 31680     | 26866     | 0.2861           | 0.3209        | 0.3025    |
| Pisces          | SNPS        | PASS          | 3045491          | 39560            | 34627     | 3010864   | 4933      | 0.0114           | 0.8753        | 0.02245   |

## Somatic Tumor-only

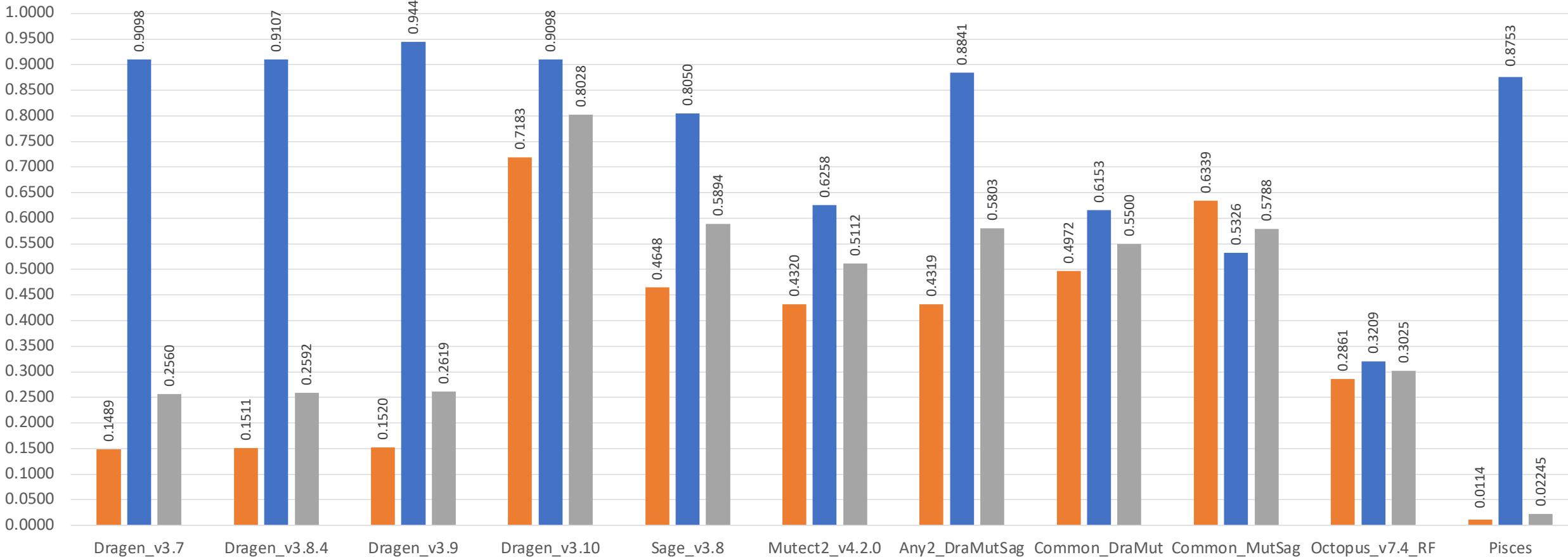
Tumor only variant calling is **performed on a tumor sample with no paired normal at the request of the research group**. This method takes advantage of the normal cell contamination that is present in most tumor samples.

| Tool                | Type        | R_Type      | Total - Q    | Total - T    | TP           | FP           | FN          | Precision     | Recall        | F1            |
|---------------------|-------------|-------------|--------------|--------------|--------------|--------------|-------------|---------------|---------------|---------------|
| Dragen_v3.7         | SNPS        | PASS        | 241631       | 39560        | 35990        | 205641       | 3570        | 0.1489        | 0.9098        | 0.2560        |
| Dragen_v3.8.4       | SNPS        | PASS        | 238448       | 39560        | 36026        | 202422       | 3534        | 0.1511        | 0.9107        | 0.2592        |
| Dragen_v3.9         | SNPS        | PASS        | 239362       | 39560        | 36387        | 202975       | 2137        | 0.1520        | 0.9445        | 0.2619        |
| <b>Dragen_v3.10</b> | <b>SNPS</b> | <b>PASS</b> | <b>50108</b> | <b>39560</b> | <b>35993</b> | <b>14115</b> | <b>3567</b> | <b>0.7183</b> | <b>0.9098</b> | <b>0.8028</b> |
| Sage_v3.8           | SNPS        | PASS        | 68504        | 39560        | 31844        | 36660        | 7716        | 0.4648        | 0.8050        | 0.5894        |
| Mutect2_v4.2.0      | SNPS        | PASS        | 57304        | 39560        | 24757        | 32547        | 14803       | 0.4320        | 0.6258        | 0.5112        |
| Any2_DraMutSag      | SNPS        | PASS        | 80974        | 39560        | 34974        | 46000        | 4586        | 0.4319        | 0.8841        | 0.5803        |
| Common_DraMut       | SNPS        | PASS        | 48954        | 39560        | 24342        | 24612        | 15218       | 0.4972        | 0.6153        | 0.5500        |
| Common_MutSag       | SNPS        | PASS        | 33237        | 39560        | 21068        | 12169        | 18492       | 0.6339        | 0.5326        | 0.5788        |
| Octopus_v7.4_RF     | SNPS        | PASS        | 44374        | 39560        | 12694        | 31680        | 26866       | 0.2861        | 0.3209        | 0.3025        |
| Pisces              | SNPS        | PASS        | 3045491      | 39560        | 34627        | 3010864      | 4933        | 0.0114        | 0.8753        | 0.02245       |

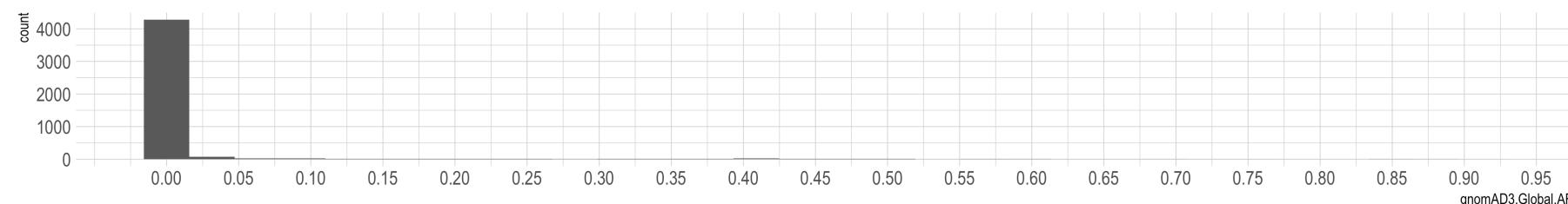
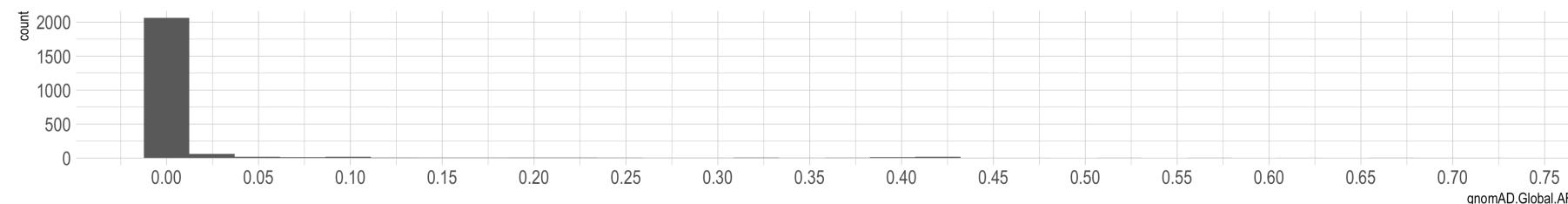
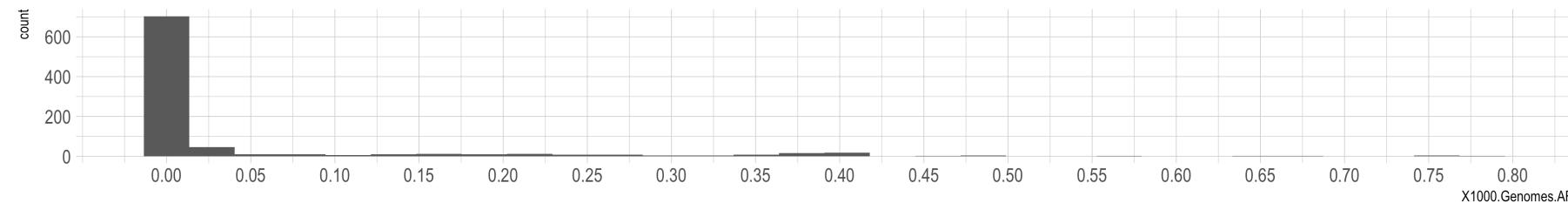
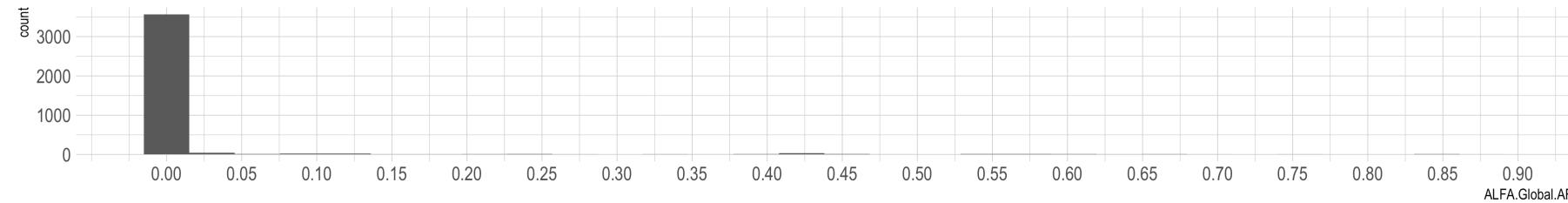
**Germline tagging on Tumor Only:** The minimum alternative allele count in population database for a variant to be defined as germline. The default=50.

# Somatic Tumor-only

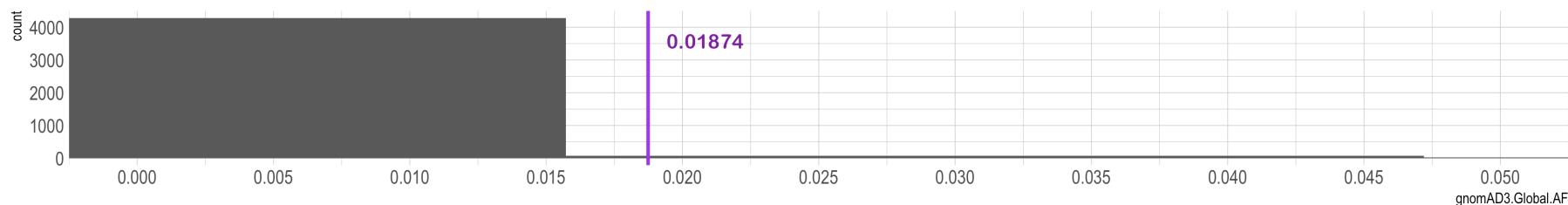
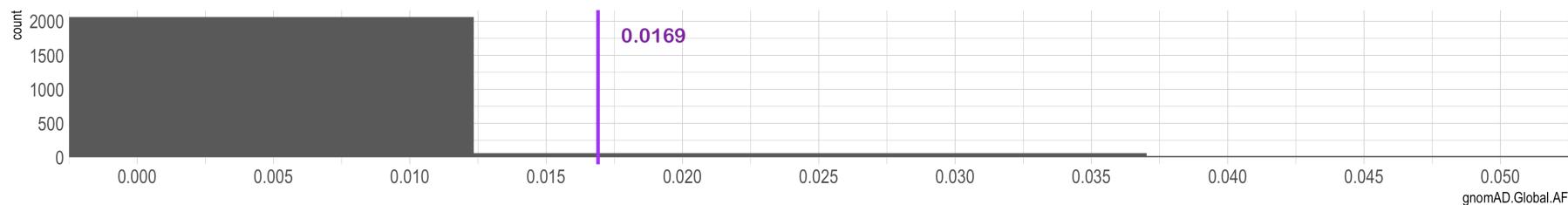
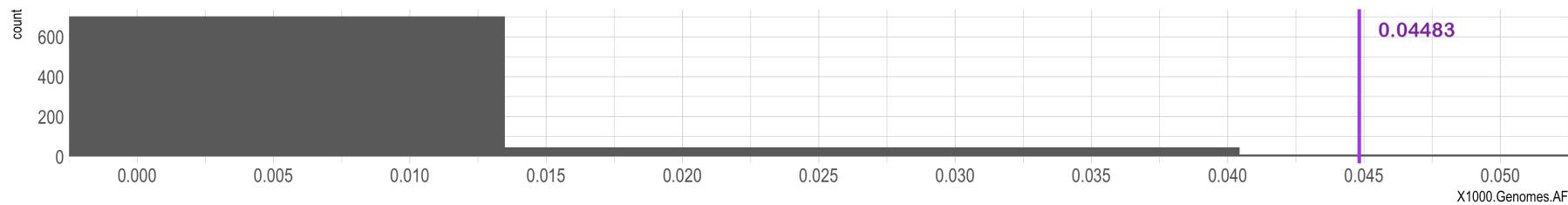
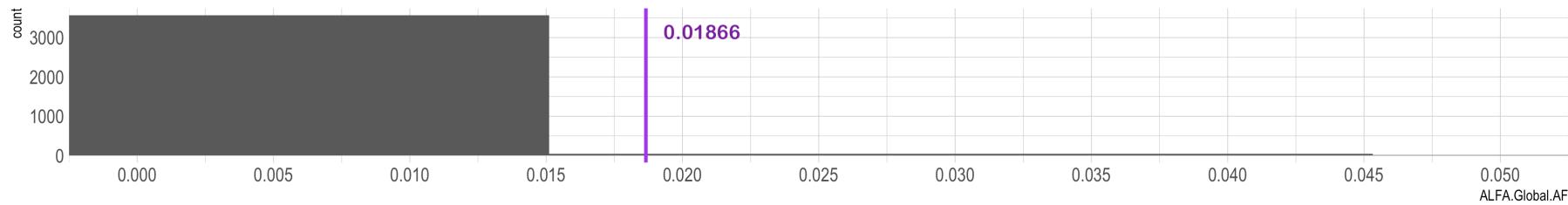
Precision    Recall    F1



# Population Frequencies in HiConf



# Population Frequencies in HiConf



# Acknowledgements

- NCI CCR-SF IFX
  - Yongmei Zhao
  - Tsaiwei Shen
- NCBR
  - Justin Lack
  - Skyler Kuhn
  - Vicky Chen
- ABCS – FNLCR